

The British Journal for the Philosophy of Science

VOLUME XII

AUGUST, 1961

No. 46

UNIVERSALITY AND NECESSITY *

WILLIAM KNEALE

IN THIS paper I wish to consider some attempts which philosophers have made to define the notion of necessity by reference to universality. I think these attempts are mistaken, and I shall put forward some arguments in justification of my view; but my main purpose is not so much to argue a case against the suggested definitions as to bring them together and to point out some similarities and differences which have not been sufficiently noticed. I think that by so doing I may perhaps arouse doubts in the minds of some who have not felt doubts on the subject before. But apart from that I shall be glad to hear a discussion of the subject, and that is my main reason for choosing it.

I

The first reductionist attempt to which I wish to draw attention is that of Hume. In his analysis of the notion of causation he remarks that this relation is supposed to involve necessary connection and asks whence the idea of such connection can be derived. The question seems puzzling to him because he has committed himself to the principle that every simple idea is derived from a simple impression and yet finds himself driven to say that no necessary connection can ever be discovered by perception between events we call causes and events we call effects. After a lot of beating about the neighbouring fields he comes in the end to the conclusion that the necessity of which we speak in this connection is merely a projection upon the world of a feeling of inevitability which we have when we pass in thought from a cause to its effect, and that this feeling of inevitability is the result of an association established by constant conjunction of instances of the first with instances of the second.

* Read to the Annual Conference of the British Society for the Philosophy of Science held at Bristol, September 1960.

Now this is not strictly speaking a definition of causal necessity in terms of universality, but it is certainly an attempt to explain away the necessity of which many philosophers have spoken in this connection. Hume does not for a moment deny that in ordinary life and in science we talk of necessary connection in nature, but he tries to explain how this can be so without admitting that there are in truth any necessary connections in nature, and he thinks that the key to an understanding of the situation is recognition of the importance of the constant conjunction of the events we call cause and effect. As though to mark the crucial importance of this notion, he prints the words 'constant conjunction' in small capitals at their first appearance.

Probably no modern philosopher would wish to defend Hume's doctrine of causation just as it stands. It is now obvious that he made a mistake in talking of antecedent causation as the only relation between events by reliance on which we can make inferences from the observed to the unobserved. It is obvious also that what he says about causation does not suffice to distinguish a cause from a sure sign of something to come. And his psychological account of the origin of our talk of causal necessity is not likely to find much favour with anyone now. But there are a great many modern philosophers who think that Hume was right in principle; and if asked to say just what it was he contributed here to philosophy, they reply that he freed the notion of natural law from a confusion with logical necessity in which it had been involved by previous philosophers. These modern followers of Hume do not, of course, wish to deny that we often talk of an event's being made necessary by another, but they think they can explain this by an adaptation of Hume's argument. Instead of saying that talk of necessity is projection of our feelings on the world, they say that it arises in connection with the use of natural laws as premisses for inference and is quite correct so long as it is understood in that connection. When I find something which is A, I may properly say that it must be B, if I am speaking to people who are already convinced that every A thing is B. For in this context the word 'must' serves only to indicate that my assertion is an inference from the premisses 'Every A thing is B, and this is an A thing.' But, they say, philosophers have sometimes mistakenly supposed that the word 'must' indicates a necessity in the law itself. Frege seems to have held this view of natural necessity when he excused himself in his *Begriffsschrift* for not dealing with modal notions. There is a very clear presentation of this view in Professor Popper's contribution to a symposium on the question, 'What can Logic do for

Philosophy?', in the *Aristotelian Society Proceedings, Supplementary Volume for 1948*. I do not think it is unfair to describe it as a modern version of Hume's theory and to say that it is an attempt to explain physical or natural necessity by reference to universality.

Historically the doctrine that there is no necessity in nature itself has often been associated with a conventionalist account of such necessities as men claim to know *a priori*, but I do not think it is quite correct to father the conventionalist thesis on Hume. When he spoke of relations between ideas and contrasted them with matters of fact, he may sometimes have meant relations of inclusion such as would be revealed by definitions; but I think that he sometimes had in mind relations of positive opposition, for example between colours, and that his philosophy contained in fact a remnant of Locke's doctrine of *a priori* knowledge as knowledge of relations between ideas. So far as I know, the first clear statement of the conventionalist thesis in conjunction with a denial of necessity in nature is to be found in Berkeley's *Philosophical Commentaries*, §§ 732-5, where he writes:

The reason why we can demonstrate so well about signs is that they are perfectly arbitrary and in our power,—made at pleasure. The obscure ambiguous term *relation*, which is said to be the largest field of knowledge, confounds us, deceives us. Let any man show me a demonstration, not verbal, that does not depend either on some false principle or at best on some principle of nature which is the effect of God's will and we know not how soon it may be changed.
Qu: What becomes of the *aeternae veritates*? *Ans:* They vanish.

In recent years the thesis has been popularised by Wittgenstein. Whether or not he intended to preach a conventionalist theory of *a priori* knowledge when he published his *Tractatus*, I do not know: there are some indications that he did not. But that was the interpretation which his disciples put on his work, and he certainly held the doctrine in his later years. Because he said that all necessity was logical necessity, what we claim to know *a priori* is often now said to be all logical, even though some of it is not obviously connected with the subject studied by Aristotle and Frege. I think this usage has misled a lot of philosophers, but there is no need to discuss it here.

So far what I have said is common knowledge. But I come now to views that are not so well known—those of Bolzano and Tarski about logical necessity. In his *Wissenschaftslehre* of 1837 Bolzano produced an account of logic as the science of sciences, which is very different from

the psychologising empiricism of Hume. Because he was a Roman Catholic priest and a professor of the philosophy of religion in Prague until deposed for expressing liberal opinions on politics, it has sometimes been supposed that he derived his characteristic views from medieval sources. He may no doubt have been influenced by reading of medieval authors, but he does not often quote them in his *Wissenschaftslehre*. In fact I have found there only three references to medieval philosophers, and those not very important. On the other hand there are many references to 'the great Leibniz', and it seems to me reasonable to think of Bolzano as working in the Leibnizian tradition, though, at the time he wrote, most of Leibniz's important work on logic was still unpublished. In various places Leibniz had pointed out that conventionalism was an unsatisfactory theory of logical necessity, since even conventions such as definitions must be shown to be consistent, and in his *Dialogue on the Connexion between Words and Things* he had ridiculed the suggestion that truth could be said to belong to sentences considered as patterns of ink. In his view, if we talk of the truth of sentences (i.e. *propositiones* in his Latin), we must consider not only actual sentences but also possible sentences, since there may well be undiscovered truths, and this introduction of talk about possible sentences shows that we are not really concerned with the speech habits of human beings but with something deeper which is presupposed in any attempt to regulate speech habits. Bolzano talks in this context of *Sätze an sich* (or propositions in the modern sense of that word). The phrase is difficult, if not impossible, to translate into English, because we have no word with precisely the same range of uses as the German *Satz*. Like our 'statement' it can mean either a form of words or what a form of words expresses, but unlike 'statement' it need not be confined to the context of assertion. Sometimes it has the sense of our 'clause' as in the grammatical term *Nebensatz*. Sometimes, on the other hand, it has the sense of 'thesis' or 'principle' as in the phrase *der zweite Satz der Wärmelehre* for 'the second law of thermodynamics'. I mention all this for two reasons, first in order to show that Bolzano did not belong to the Berkeley-Hume-Mill tradition but rather to what one might call the tradition of logical realism of which Leibniz and Frege are the greatest representatives, and secondly in order to explain his terminology.

According to Bolzano a proposition (i.e. a *Satz an sich*) is analytic with respect to a certain constituent if the class of propositions that we can obtain by substitution for that constituent (including here under

substitution' replacement of the original constituent) consists entirely of true propositions or entirely of false propositions. Bolzano himself admits that the notion of analytic proposition which he has defined in this way is much wider than that of Kant; and he says that, if the part of an analytic proposition which is considered invariant (i.e. not open to substitution) contains only logical notions, it may perhaps be useful to describe the whole proposition as logically analytic, or analytic in the stricter sense (*Wissenschaftslehre*, §§ 147-148). Apparently he thinks that this is something like the sense intended by Kant; but he does not attach much importance to it, because he does not think it possible to draw a clear line between logical and non-logical notions.

There are a number of curious features in this passage of Bolzano's work. The first is a purely verbal point. Bolzano has so defined the word 'analytic' that it can be applied to false propositions. This is contrary to Kant's usage, but it is easy to see why Bolzano thought it reasonable. He wanted to use the words 'analytic' and 'synthetic' as exhaustive of the realm of propositions, and he noticed that according to Kantian usage self-contradictory propositions would be neither analytic nor synthetic. The practice of later philosophers has followed Kant's lead rather than Bolzano's and it seems that we must either reconcile ourselves to saying that 'analytic' and 'synthetic' are not contradictory opposites or try to secure the exhaustiveness of the division by applying it only to truths. Secondly, Bolzano's account of analytic propositions seems to be based on the very naive assumption that a proposition must contain distinguishable constituents corresponding to all the distinguishable constituents of a sentence that expresses it, and *vice versa*. I shall say no more about this. Thirdly, Bolzano has given a definition of 'analytic' such that a proposition can be analytic by virtue of natural laws or even by virtue of mere accidents. Consider for example the sentence 'Kant was not an eighteenth-century philosopher who died on the anniversary of his birth.' If it so happens that no eighteenth-century philosopher died on the anniversary of his birth, all the sentences that can be made from this by substitution of other names for 'Kant' express true propositions, and so the proposition expressed by the original sentence is analytic with respect to the constituent Kant according to Bolzano's definition.

The most interesting and valuable element in Bolzano's discussion of the distinction between analytic and synthetic propositions is his recognition of the fact that there are certain characters which belong to propositions in virtue of their structure. According to his way of

talking a proposition may be universally valid (*allgemeingültig*) and so analytic with regard to substitution for certain constituents though not with respect to substitution for certain others. This is intelligible, but rather confusing, and it seems preferable to say (as later logicians have done) that universal validity (or validity for short) belongs primarily to certain propositional patterns and only secondarily to the propositions which exemplify them. Similar paraphrases can be applied also to what Bolzano says of compatibility and derivability.

According to Bolzano's own way of describing the matter, the propositions M, N, O, . . . follow or are derivable (*ableitbar*) from the propositions A, B, C, D, . . . with regard to the constituents i, j, . . . if any set of ideas which yield a set of true propositions when substituted for i, j, . . . in A, B, C, D . . . do the same when substituted for i, j . . . in M, N, O, . . . (*Wissenschaftslehre*, § 154). In other words this means that a proposition called a conclusion is entailed by a set of propositions called premisses if the argument constituted by the association of the premisses and the conclusion exemplifies a pattern of argument for which every exemplification with true premisses also has a true conclusion. For a reason which I shall notice later, it is not now considered correct to identify following-from with being-derivable-from, as Bolzano does, but without going into details of modern discoveries we can see that there is something wrong in the use of the word 'derivability' for the relation defined by Bolzano. For a proposition cannot properly be said to be derivable from a set of premisses unless it is possible to establish that if the premisses are true the proposition is also true without first establishing whether or not the premisses and the proposition are true. But the relation defined by Bolzano might hold when this condition was not fulfilled. Just as according to his definitions a proposition can be analytically true by accident, so too one proposition may follow from another by accident, that is to say in such a way that the truth of the universal proposition about the results of substitution can be known only by an examination of the individual results. This can be seen most easily from consideration of the limiting case where the proposition called a consequence has no constituents in common with any of the premisses. For then, according to Bolzano's definition, its being a consequence of the premisses depends solely on its being true. He does not make this point explicitly in his section on the relation of being a consequence, but in his previous section on compatibility he says that a false proposition which contains none of the constituents for which substitution may be

made in a certain set of propositions is incompatible with those propositions (*Wissenschaftslehre*, § 154), and according to his own account of the connection between the two sections this is just another way of saying what I have just said about being a consequence. Perhaps Bolzano overlooks it because he wants to think of the relation of being a consequence as holding always in virtue of a general rule. For he says in a note at the end of this section that when Aristotle uses the phrase *συμβαίνει ἐξ ἀνάγκης* ('it follows of necessity') to describe the relation of the conclusion to the premisses in a valid syllogism, even though premisses and conclusion may be alike false, he must surely mean that every argument of the form exemplified leads to a true conclusion if only the premisses are true.

All this might be dismissed as of no more than antiquarian interest were it not for the fact that Bolzano's theory has been restated independently in our time by Tarski. In his article 'On the Concept of Logical Consequence' (first published in Polish in 1936 and republished in English in *Logic, Semantics, Metamathematics*, p. 409) he rejected the Wittgensteinian account of logical necessity as unclear and said that Gödel's discoveries had shown the need for a new account of the relation of consequence:

In order to obtain the proper concept of consequence, which is close in essentials to the common concept, we must resort to quite different methods and apply quite different conceptual apparatus in defining it.

The method which he then proposed was in essentials that of Bolzano, though at the time of his first publication on the subject he knew nothing of this part of Bolzano's work.

Tarski himself has summarised his account of the consequence relation as follows:

Let L be any class of sentences. We replace all extra-logical constants which occur in the sentences belonging to L by corresponding variables, like constants being replaced by like variables, and unlike by unlike. In this way we obtain a class L' of sentential functions. An arbitrary sequence of objects which satisfies every sentential function of the class L' will be called a *model* or *realization of the class L of sentences*. . . . If in particular the class L consists of a single sentence X , we shall also call the model of the class L the *model of the sentence X* . In terms of these concepts we can define the concept of logical consequence as follows:

The sentence X follows logically from the sentences of the class K if, and only if, every model of the class K is also a model of the sentence X .

Later he adds:

We can agree to call a class of sentences *contradictory* if it possesses no model. Analogously a class of sentences can be called *analytical* if every sequence of objects is a model of it. Both of these concepts can be related not only to classes of sentences but also to single sentences.

Tarski himself insists that his essay is an attempt to work out clearly the implications of the old doctrine that the relation of logical consequence holds between statements in virtue of their forms. But his development of the doctrine has some curious features. I wish to draw attention to one only, and for this purpose I shall quote again from his own work. At the end he writes:

Underlying our whole construction is the division of the terms of the language discussed into logical and extra-logical. This division is certainly not arbitrary. . . . If, for example, we were to include among the extra-logical signs the implication sign, or the universal quantifier, then our definition of the concept of consequence would lead to results which obviously contradict ordinary usage. On the other hand, no objective grounds are known to me which permit us to draw a sharp boundary between the two groups of terms. It seems to be possible to include among logical terms some which are usually regarded by logicians as extra-logical without running into consequences which stand in sharp contrast to ordinary usage. In the extreme case we could regard all terms of the language as logical. The concept of *formal* consequence would then coincide with that of *material* consequence.

By 'material consequence' he evidently means here the converse of what logicians call material implication. We may add that in this extreme case the concept of the analytic would coincide with that of the true.

In effect Tarski expresses the same doubts as Bolzano about the possibility of drawing a clear line between the logical and the extra-logical; and having made this point he rightly goes on to show that by placing the boundary between formal and material in different places we can pass from the ordinary notion of logical consequence to much weaker notions. For just as many philosophers who profess to follow Hume have said that the necessity with which an effect follows its cause is no more than the inclusion of this particular sequence in a constant natural association, so he says that the necessity with which a consequence follows in the logical sense of 'follows' from premisses is no more than the inclusion of this particular sequence in a universal fact about the satisfaction of certain sentential functions. If, following his own suggestion, we enlarge the range of sentential functions under consideration by including among the constant factors of our sentences not only the traditional logical constants but all unrestricted

general terms such as 'man', 'iron', 'fire', etc., we find that his definition of consequence covers not only logical consequence but also all consequences in virtue of natural laws. In short, his account of necessity amounts to a generalisation of the constancy theory of natural necessity. But the result can give no pleasure to followers of Hume, since it involves rejection of their view that there is a fundamental difference of kind between logical and natural necessity.

In defence of Tarski's definition it may perhaps be argued that for a sentential function such as 'If p then not-not- p ' we can know *a priori* that it is satisfied by every sequence of objects, whereas for a sentential function such as 'If x has taken arsenic x will die' we can only conjecture universal satisfaction on empirical grounds. This is true, but there is nothing in Tarski's account of the matter to explain the difference, and he does not allude to it himself. Nor is it appropriate that a distinction of kinds of consequence should be made to depend on a distinction between cases in which we can and cases in which we cannot gain knowledge *a priori*. On the contrary, the epistemological distinction should be explained by an account of the difference of the cases; and it is just this which is lacking so far.

3

For my own part I do not agree with the project of trying to define necessity by reference to universality, and I have argued against even the more popular part of the project, namely that of accounting for natural necessity by a modernised version of Hume's analysis. In my *Probability and Induction* of 1949 I pointed out that when we enunciate a natural law, or what we suppose to be such, in the form 'Every F thing is G', we do not think of it as a merely *de facto* generalisation of the kind logicians express by writing ' $(x) [Fx \supset Gx.]$ '. Although we may not use any modal word such as 'must' or 'necessarily', we assume that our pronouncement commits us not only to asserting that everything which actually has been or will be F has been or will be G but also to asserting that if anything which is not as a matter of fact F were F it would also be G. This, I thought, was sufficient to show the inadequacy of a Humean account of natural law. The same point has been made by a number of other philosophers, but the conclusion which I have drawn has not been accepted by all, or indeed by many. Although it seems to be generally agreed that subjunctive conditional statements can be inferred from statements of natural law, and that

statements of natural law must for this reason be distinguishable from statements of accidental universality such as we find in works of history and geography, for example 'All mountains in the United Kingdom are of less than 5,000 feet in height', it is sometimes thought that this feature of statements of law can be explained without supposing that they involve any more than Hume allowed.

Already in 1948 Professor Popper argued for this view in the symposium 'What can Logic do for Philosophy?' and he defended it again in 1949 in 'A Note on Natural Laws and so-called "Contrary-to-fact" Conditionals' which he contributed to *Mind*. His thesis was that the difficulties of persons like myself arose from failure to notice the difference between terms which can be defined extensionally and those which cannot be so defined. It is true, he said, that 'All my friends speak French' does not entail 'If Confucius were one of my friends he would speak French', but that is because anyone who utters the first statement is thinking of the class of his friends as closed, whereas anyone who utters the second statement is thinking of the class of his friends as open. In other words, the expression 'my friends' is not used in quite the same way in the two statements. When, however, it is maintained that sentences which purport to state natural laws are equivalent to universal material implications, it is to be understood that the terms involved are not mere substitutes for lists of proper names but unrestricted general descriptions. And once this is conceded, there should in his opinion be no difficulty about the derivation of contrary-to-fact conditionals.

Against this I argued in a paper contributed to *Analysis* in 1950 (later reprinted in Miss Margaret Macdonald's collection *Philosophy and Analysis*) that Popper's theory would not do what he wanted of it because it still allowed for no distinction between laws and merely accidental generalities. Philosophers who treat suggestions of law as universal material implications say in effect that there is no sense in talking of historical accidents on the cosmic scale. According to their account of the matter, when we consider the hypothesis of a connection between any two characters expressed by unrestricted general descriptions, there are only two possibilities with which we have to reckon. Either it is a law of nature that every A thing is B, or there has been or will be somewhere at some time an A thing that was not or is not B. In short, they are committed to the view that every natural possibility (i.e. every state of affairs not excluded by a law of nature) must be realised somewhere at some time. But this is certainly not

what we ordinarily think, and any philosophical theory which leads to this conclusion should be regarded with great suspicion.

In an appendix to his *Logic of Scientific Discovery* of 1959, Popper concedes that something more must be done to distinguish laws from accidental generalities and produces the following definition of natural necessity:

A statement may be said to be naturally or physically necessary if, and only if, it is deducible from a statement function which is satisfied in all worlds that differ from our world, if at all, only with respect to initial conditions (p. 433).

As he goes on to remark, this proposed definition makes all laws of nature, together with all their logical consequences, naturally or physically necessary, but excludes from that status generalisations which hold merely because of the *de facto* arrangements of things in the world we know. Personally I find it acceptable. For to say that a statement function is satisfied in all worlds that differ from the actual world, if at all, only with respect to initial conditions is to say in effect that it holds for all *possible* worlds that contain instances of the same attributes and relations as are exemplified in the actual world and of these only; and what holds for all possible worlds is obviously necessary. I agree, of course, with Popper that we cannot know for certain whether a generalisation which we put forward is in truth a law of nature with natural necessity as he has defined it; but to say this is only to admit that we cannot establish laws of nature *a priori*. Obviously experience, which is always of the actual world, cannot guarantee a generalisation which is supposed to hold for worlds other than the actual world, and those who speak, as I have done, of natural necessity do not wish to say that it can. The important thing in Popper's new definition, and what makes it acceptable to me, is just that it connects the notion of natural law with that of validity for states of affairs other than the actual. Unfortunately, however, Popper himself seems to be in some confusion about the effect of his concession. Perhaps I have misunderstood him. If so, I hope to be enlightened. But some passages which I am going to quote seem to me very puzzling indeed.

Just before his new definition of natural necessity Popper writes:

As Tarski has shown, it is possible to explain *logical necessity* in terms of universality: a statement may be said to be logically necessary if and only if it is deducible (for example by particularisation) from a 'universally valid' statement function, that is to say, from a statement function that is *satisfied by every model* (this means true in all possible worlds). I think we may explain by the same

method what we mean by *natural necessity*; for we may adopt the following definition (*L.S.D.* p. 432).

But a few pages later he says:

I regard, unlike Kneale, 'necessary' as a mere word, as a label for distinguishing the universality of laws from 'accidental universality'. Of course any other label would do as well, for there is not much connection here with logical necessity. I largely agree with the spirit of Wittgenstein's paraphrase of Hume: 'A necessity for one thing to happen because another has happened does not exist. There is only logical necessity' (*L.S.D.* p. 438).

As they stand, these two passages are inconsistent. For the first recommends a new account of natural necessity by saying that it is framed after the pattern of Tarski's account of logical necessity, whereas the second says there is no important connection between natural necessity and logical necessity. And the first implies acceptance of Tarski's account of logical necessity, whereas the second implies acceptance of Wittgenstein's account of logical necessity, which Tarski himself, as we have seen, found unsatisfactory. I do not wish, however, to dwell on the second passage, and I quote it here only in order to indicate that there seems to be some uncertainty in Popper's own mind about his commitments. It is the first passage which interests me, and it is this I wish to examine.

At the beginning Popper says that Tarski has shown that it is possible to explain logical necessity in terms of universality. I think this is at any rate a correct report of Tarski's intention. But, as I have already remarked, neither Bolzano nor Tarski ever produced a delimitation of the logical realm which satisfied him, and there is something unsatisfactory in a theory of logical necessity which does not even explain why logic is an *a priori* science. It is obvious, of course, that truths of logic are either themselves universal in the way explained by Bolzano and Tarski or cases falling under generalisations of the kind those authors discuss. That has been taken for granted by all the great logicians, Aristotle, Chrysippus, Leibniz, Boole, Frege, and Russell. But it is not so obvious that there is nothing more to be said about logical necessity.

Towards the end of the passage I have quoted, Popper tries to make the theory more acceptable by adding in brackets 'this means true in all possible worlds'. This phrase is supposed to explain the immediately preceding phrase 'satisfied by every model', but according to Tarski's own explanation a model is an arbitrary sequence of objects which satisfies a statement function or class of statement functions, and an

analytic statement is one for which every sequence of objects is a model, i.e. one satisfied by every sequence of objects, or as we may say in another terminology, one which turns out true for every interpretation of its extra-logical signs. There is nothing here about possible worlds, and it seems clear to me that, so far as Tarski is concerned, the truth of a formula under all interpretations is just a fact about the actual world, though this actual world may be taken to include not only physical objects but an infinity of sets belonging to various levels. Indeed, if Tarski's programme is what Popper says at the beginning of the paragraph, namely to explain logical necessity in terms of universality, he has no right to speak of possible worlds at all. It was no doubt correct for Leibniz to say that necessary truths are true in all possible worlds, and even to offer this as a definition of necessity; but he did not think that he was explaining necessity in terms of universality and so showing the superfluity of modal expressions. For him the definition was merely an explanation of one modal notion in terms of another which some people find easier to grasp. If Tarski were willing to work with the notion of possibility as Popper supposes, he could define the notion of consequence much more easily than he has done, by saying simply that a statement is a consequence of a class of premisses if the conjunction of the premisses with the negation of the statement does not represent any possible state of affairs. But I feel sure that he would think this a less satisfactory definition than that he has given, precisely because it contains a modal word.

Immediately after the sentence about possible worlds Popper goes on to say 'I think we may explain by the same method what we mean by natural necessity', and the explanation that he gives is indeed an explanation of the Leibnizian type. For it involves generalisation over all worlds that differ from our world, if at all, only in initial conditions, and these must clearly be naturally possible worlds with instances of the same attributes and relations as we find exemplified in our actual world. I hold therefore that Popper is mistaken if he thinks, as apparently he does, that he has succeeded in improving on the Humean account of natural laws while remaining true to the principle that necessity can be explained without remainder in terms of universality. In fact he has only defined one modal notion in terms of another; but he has hidden this from himself for the moment by talking of all worlds which differ from our world, if at all, only in initial conditions.

In my opinion, it is a mistake to try to explain away the modal notions. Instead of trying to reduce necessity to universality we should,

I think, take the notion of necessitation as fundamental and say that the logical constants, about whose recognition Bolzano and Tarski were both puzzled, are just those signs which can be defined without remainder by formulation of principles of necessitation. Popper himself has taken this line in a number of papers written about 1947-8 (in particular two papers 'On the Theory of Deduction' which he contributed to the *Proceedings of the Royal Netherlands Academy of Sciences* in 1948), and I have tried to follow him in a paper called 'The Province of Logic' which I contributed to *Contemporary British Philosophy, Third Series*. What I suggest is simply that we should regard formal logic as the pure theory of necessitation, that is as the study of what can be said about necessitation in general without regard to those special principles of necessitation which hold for various subject matters. Although he has come very near to saying this, Popper has in the end drawn back and professed himself a follower of Hume and Wittgenstein, that is, an upholder of the anti-scientific doctrine of conventionalism which was first promulgated in modern times by Berkeley. His reason for taking the line which he does is apparently that he cannot understand what a non-formal principle of necessitation would be and suspects that admission of any such would commit him to a terrible evil called essentialism. To this I reply that a non-formal principle of necessitation would be exactly what he has allowed a law of nature to be, namely a generalisation which holds for all possible worlds of some kind, and that it is highly paradoxical to suggest, as he does in a passage which I quoted, that there is none but an arbitrary linguistic connection between natural and logical necessity.

University of Oxford

ON THE RELATIONSHIP BETWEEN METHODOLOGY IN SCIENTIFIC RESEARCH AND THE CONTENT OF SCIENTIFIC KNOWLEDGE *

DAVID BOHM

SCIENCE is characterised in an essential way by what is commonly called the *scientific method*. The main content of this method is a set of general rules, criteria, and directives, partially implicit and partially explicit, as to how to set about doing scientific research, to evaluate the results obtained in doing so, and to modify one's subsequent work on the basis of such evaluations. For example, one should do experiments, set up hypotheses to explain their results, test these hypotheses by experiment, discard those that are not confirmed, set up new hypotheses to explain the new experiments, etc.

Scientists generally apply the scientific method, more or less intuitively, and can usually tell in practice whether they are following it properly or not. Nevertheless, the question of proper method becomes ambiguous under certain circumstances. This happens especially in the development of new sciences, and in fields that are on the border line between science and something else (e.g., psychology, sociology, etc.). Even in traditionally scientific fields, however, the question of method sometimes becomes unclear. For example, in the growth of relativity and quantum theory a great deal of attention had to be focused on the meaning of measurements and on the proper rôle of theories (to such an extent, for example, that ether theories were discarded partly for the reason that they were criticised as 'metaphysical' because the velocity of the ether was not capable of being measured).

The reason why the intuitive application of the scientific method becomes ambiguous in borderline regions is basically that all the terms which appear in its rules, criteria, and directives depend for their clear meaning on what may be called a general framework of concepts, ideas, procedure, etc., which function together as a coherent whole. Against a background in which this general framework applies, one can consider specific problems which do not call the framework into

* Read to the Annual Conference of the British Society for the Philosophy of Science at Bristol University in September 1960.

question. Within this limited domain, it is clear how one is to go about finding the facts, setting up hypotheses, testing them, etc. But as we approach the edge of such a domain, these questions become less clear.

In order to bring out this point in more detail, let us consider briefly Professor Popper's well known thesis, that whereas a hypothesis claiming to have universal validity cannot be verified by any number of confirmations, it can be refuted by one falsification. In the domain where the general framework of a science is clear and well-formed, it is easy to see what constitutes a falsification. For example, it is obvious that the hypothesis that an object in a vacuum will fall at a constant velocity can be tested by an experiment, aimed at either falsifying or confirming it. This is because a very large number of things can already be accepted. Thus, one begins with the knowledge that there is a well defined frame of space and time, that an object occupies a definite position at a definite time, that its position can be measured in certain ways, etc. Within this framework, it is possible, so to speak, to ask of Nature a question which has a definite 'yes' or 'no' answer; for example, 'Does or does not the object occupy such and such a series of positions at such and such a series of times?'

However, as we approach the boundary of the domain where a given framework applies, then it is no longer clear what is the right question to ask. For example, in pre-relativistic theory, it was always valid to ask 'Did event A come before event B or after?' Relativity theory showed, however, that if events are outside each other's light cones, there is no unique answer to this question. The answer depends, in fact, on the speed of the frame in which space and time are measured. This is because space and time are different in properties from what was assumed in the Newtonian conception. Similarly, quantum theory showed that in the micro-domain, one can no longer answer 'yes' or 'no' to the question of whether there is an electron at a given point or not. This is because an electron is not just a particle, but has wave-like properties also. Examples of this kind can be found in a wide variety of fields.

It is clear then that underlying the problem of falsification and confirmation, there is something very much more fundamental; namely, that of framing questions that have a definite 'yes' or 'no' answer in a given situation. It is common in everyday experience that most questions cannot be answered with a clear 'yes' or 'no'. It is sometimes overlooked, however, that the same is generally true in science, and that a great deal of work is needed to arrive at a question that does

have such an answer. In fact, it is frequently realised that half the battle is over when we know what are the right questions to ask.

What must be emphasised here is that the *form* in which a question is put constitutes an implicit hypothesis about the object of the question; namely, that the object is such that this question has a clear 'yes' or 'no' answer. The classical example of such a question is 'Have you stopped beating your wife?' Evidently, this question implies that you are the sort of person who beats his wife. Similarly, the question 'Is the electron at a given point or not?' implies that the electron is the sort of thing that can be at a given point. Most scientific questions are in some way similar, in the sense that whether the answer is 'yes' or 'no' is much less important than whether they do, in fact, have definite answers of this kind.

It is clear that an essential feature of the scientific method is that it seeks hypotheses leading to questions that can be answered definitely 'yes' or 'no' by experiments or observations. But since we recognise that a science may pass through phases (especially while it is developing rapidly) in which the proper questions having this character are not even known, we are led into a whole new series of problems relating to the rôle of concepts, theories, and hypotheses in scientific research. Among these, one of the most important is that of what can be meant by the term 'fact'. Now, as is well known, a fact is a contingency, in the sense that *a priori* it could either be so or not so. As a result, some empirical criteria are needed to decide in each case whether it is so or not. If, however, we are working in a domain in which the usual questions that have been asked of Nature do not have a clear 'yes' or 'no' answer, then our customary criteria of factuality become ambiguous. And where these criteria are ambiguous, we are not really dealing with facts, even though we may at first sight think that we are.

It is not at all uncommon for a scientist to be mistaken as to whether something which he takes to be a fact can really even be one. For example, a physicist not aware of the theory of relativity could say 'A possible fact is that events A and B are simultaneous for all observers'. Knowing the theory of relativity he would only say 'A possible fact is that they are simultaneous for an observer moving at a given velocity'. In terms of the quantum theory he could go further to say 'There is an irreducible degree of ambiguity in the relative time of two events (implied by the uncertainty principle), so that it may not even be a possible fact for two events to be simultaneous for a single observer.'

Thus, guided by different conceptions, one is led to seek different kinds of facts, some of which may be possible in a given field, and some not.

It follows from the above that facts are not like things that exist in nature independently of man, so that they can, so to speak, be gathered as if they were stones or flowers. Rather, as the derivation of the word indicates, they are *made* (or manufactured). In this regard, nature may be compared to the raw material, while the fact is a finished product, having a form determined in part by man. Like a pair of shoes, for example, the fact reflects both the material (i.e. nature) and the process by which it was made. However, even the form is not wholly at man's disposal, since not every material will sustain every form (e.g. shoes cannot be made out of water). Similarly, the content of each field of scientific research calls for facts having appropriate forms. As we enter new fields, we must generally give new forms to the facts, forms that will be appropriate to the character of nature in the new domain.

It is in the form of the facts that the whole body of the existing framework of concepts and laws plays a crucial rôle. For, as we have already seen in a number of examples, each such framework suggests certain questions which should have a 'yes' or 'no' answer. In the domain where a given framework is valid, all facts can ultimately, in principle, be expressed by giving such an answer to an appropriate set of questions. In this regard, even numerical facts have such a character. Thus every real quantitative measurement (which is necessarily of finite accuracy), answers the question 'Is a given parameter between certain prescribed limits or not?'

But, of course, it is not enough merely to list the facts. It is also necessary in the scientific method that the fact should be relevant. It is here that laws play a fundamental part. For, in scientific research the relevance of facts to each other generally consists in their being related in some coherent system of law. Indeed, it is a basic assumption underlying the scientific method that any set of facts can, in principle, ultimately find their places in a suitable system of law. Of course, it is not expected that this will ever really be accomplished in such a way as to include the totality of all possible facts. Therefore, this principle, like that of seeking questions having a definite 'yes' or 'no' answer, plays the rôle of a directive, showing in what way we should aim to develop our ideas.

At any given moment then, we are faced, not with one coherent system of facts, but rather with a large number of partial systems, each

having some intermediate degree of coherence. (There are different fields of study, different domains and levels in each field, etc.). Within each partial system, certain kinds of facts are possible. For example, in the context of everyday experience, there is an enormous variety of facts (e.g. a bell rings, the sun shines, there is a speck of dust here, a cloud there, etc.). It is recognised that most of these facts are not relevant to each other, or to the expression of the laws of physics, chemistry, biology, etc. (except perhaps in an extremely remote way). In order to obtain facts that are relevant in a given field, we need a set of concepts and a corresponding set of laws, which enables us to express the facts *as being in the field in question*. Thus, to study classical dynamics, it is no use to consider the colour or texture of the objects (which are facts in the context of everyday experience). Rather the facts and laws must be expressed in terms of co-ordinates and velocities of various bodies. On the other hand, in quantum mechanics the facts and laws are expressed in a very different way, namely, in terms of energy levels, cross-sections, probabilities, wavelengths, parities, etc. Indeed, the whole methodology of modern physics has been developed largely in the effort to 'manufacture' facts of this general kind.

The laws of a particular science not only play their parts in a general scheme with the aid of which facts can be organised into a single body of knowledge; they also help to define new forms of facts that are relevant for the science in question. For each law implies a whole set (generally inexhaustible) of relationships in the totality of facts. *Each such relationship is itself a possible fact*, in the sense that whether it is present or not is a 'yes' or 'no' question, subject to experimental or observational test. Moreover, if a large number of different kinds of relationships is consistently found to be as implied by a certain law, then the applicability of that law is likewise taken to be a fact, at least for some domain, and to a suitable degree of approximation. Even if the law should be refuted in some other domain, or in experiments carried out to a higher accuracy, the fact of its validity in the original domain and degree of approximation is not thereby altered (e.g. as the physicist accepts as a fact the behaviour of large-scale apparatus approximately according to the laws of classical mechanics, although he uses this apparatus in studying the laws of quantum mechanics, with the aid of which he will refute the notion that the classical theory is exactly and universally valid.)

It is evident that the part played in scientific method by already known laws will also eventually be played by laws which are not yet

known, but which will come to be known in the future. But a hypothesis is just a presupposition concerning such laws, a presupposition that when it is first made has yet to be confirmed or refuted. Since the scientific method implies that we should proceed *as if* the hypothesis were true, unless and until it is actually refuted, it follows that even a hypothesis can help organise existing facts into a coherent whole and determine new forms in which relevant facts can further be developed. For, if the hypothesis is correct, it leads us to ask appropriate questions of Nature. Consider, for example, the hypothesis of non-conservation of parity, which opened up for consideration a whole new set of possible facts relating to the properties of systems under space-time reflections, properties that would not have had any meaning in the framework of previous hypotheses involving conservation of parity. Even when the hypothesis is not correct, but just on the right general lines, it may, as has frequently happened in the history of science, suggest problems that stimulate fruitful research and eventually lead to better theories. On the other hand, there is, of course, the possibility also illustrated by many examples in the history of science, that wrong hypotheses may impede research, by drawing attention away from questions that are relevant. Thus, the organisation of facts and the determination of their forms with the aid of hypotheses has both great advantages and great dangers.

The form of facts is determined, not only by concepts, laws, and hypotheses, but also by certain still more general and pervasive assumptions, which are usually implicit in what I may call habits of thought that have become associated with scientific thinking. Two examples will be considered here; first, the habit of thinking in terms of prediction, and second, the habit of specialisation.

The tendency to focus on the predictive aspects of science became important with the advent of Newtonian mechanics and its application to ballistics, the movement of planets, etc. In this branch of science, facts generally took the form, as we have already pointed out, of positions and momenta of bodies at one or more instants of time, while hypotheses were made on the general laws of motion from which one could deduce positions and momenta at other times. The test of hypotheses came to be the matching of future observations with predictions. Gradually there arose the notion that the ideal toward which physics should aim is complete predictability (as exemplified by the Laplacian demon, who could calculate the whole future of the universe if he knew the initial conditions of all the bodies in it).

METHODOLOGY AND SCIENTIFIC KNOWLEDGE

With the advent of the quantum theory, it became clear that perfect prediction is impossible, if only because at the micro-level, entities such as electrons are not the sort of things that can have, simultaneously, a well defined position and a well defined momentum. Detailed predictive laws were replaced, as is well known, by statistical laws. Nevertheless, a strong effort was made to keep the *form* of facts, laws, and hypotheses as close as possible to that of the classical scheme. Thus, it was supposed that an electron is the sort of thing that can, at each moment, have a *quantum state*, determined by its wave function, $\psi(x, t)$. Some initial measurement gives us the quantum state at a given time, t_0 , represented by the wave function $\psi(x, t_0)$. The law then takes the form of an equation (e.g. Schrödinger's equation or Dirac's equation) which predicts the future quantum state, and from which one can calculate future probabilities (i.e. $P(x) = \psi^*(x) \psi(x)$). In principle, one retained the ideal of determining the quantum state of the system at a given time, and of predicting the future of this state. As with classical mechanics, such predictions then afford a test of the theory.

It is clear that because a certain part of the general classical scheme for expressing facts, laws, and hypotheses has been retained, one still tends to regard prediction as the essence of physical theory. But is this really so? In order to define the question more sharply, we may ask whether it is essential that we first obtain our facts on the basis of a measurement of initial conditions, *then* make our calculations on the basis of laws or hypotheses, and *finally* make observations to check these calculations. Would it not be equally valid if we were presented all at once with a set of data taken at different times on the same system? Then we could consider the data corresponding to one time, t_1 , calculate what to expect for a later time, t_2 , and compare with the actual data. Surely such a test is just as good as one in which we do our calculations before we obtain the data for the later time, t_2 . Vice versa, we could also take the data for t_2 and *retrodict* the conditions at t_1 . This too would be just as good a check on the hypotheses. But equally well, we might consider 'mixed data' taken from two times, t_1 and t_2 , and see whether the remainder of the data taken either earlier, later, or in times that are between t_1 and t_2 , are connected in the way implied by the hypotheses that are being tested.

If one reflects a bit, one will see that in modern high energy physics, most of the data is of the type described above, in which prediction is essentially irrelevant. For example, in cosmic ray research, a high

energy particle may come in (at a speed close to that of light) and activate a counter, C_1 . It then passes through a block of lead, several particles emerge, and a number of counters, C_2 , C_3 , C_4 , etc., on the other side of the block of lead are activated. It would not only be unnecessary but impossible in practice for the physicist quickly to become cognizant that the counter, C_1 , had been activated, then immediately to calculate what was going to happen while the particle was in the process of passing through the block of lead, and then to verify his predictions, by seeing whether the counters C_2 , C_3 , C_4 , etc., clicked or not. What happens instead is that he is presented with a final result, in which connected events occurring at a series of two or more times are all recorded together, and he studies the relationships among these events to see whether they are as his theory says they should be.

It is clear then that the stress on prediction of future events in physics is not, in general, really justifiable. For we are frequently dealing with systems (especially with quantum mechanics and high energy physics) which are unpredictable and in which prediction is not the essential point. What is essential, quite generally, is that our theory should furnish us with a correct knowledge of the relationships in the system under investigation.¹

The habit of thought by which one identifies all lawful relationships with predictability is capable of having an important limiting influence on the directions in which scientific research can develop. For it tends to blind us to the possibility that there may be real relationships in time having no fundamental connection with predictions. As a result, there may be a large range of facts that we simply do not look for, because there is no room in our general framework of thinking to express these facts.

In order to exemplify this point in more detail, let us return to our discussion of research on high energy particles. In certain cases, where such particles from outer space are incident on a photographic plate, there is produced a very long and highly ramified chain of particles, which is called a shower. Such a shower results from a series of collisions with the atoms of the plate, in each of which two or more

¹ H. L. Armstrong, in a private communication, has made the very pertinent comparison of the rôle of predictions in scientific research to that of examinations in university studies. In a cursory view, it might seem that students frequently act as if the main purpose of their work is to pass examinations. A more careful consideration shows, however, that the purpose of the examinations is to test for how well the student knows the subject. Similarly, predictions help us to test for how well we know the relationships in Nature that we are investigating.

particles emerge. Although events are produced in a series of collisions, as described above, we are, of course, presented with the whole of the data all at once, in the form of a set of tracks which the particles leave in the photographic plate.

Now, according to current views on predictive laws, we would use the quantum mechanics to predict the probability of the first collision. We would then consider a wave function representing the actual state of those particles that emerged from this collision, and with this, we would predict the probability of the next collision. A similar procedure would then be carried out on these particles, and in this way, we could eventually discuss the development of the whole shower in a chain of collisions.

In such a treatment, the only facts that we could take into account would be statistical correlations, such as that *A* is followed by *B* in a certain fraction of the cases, *B* by *C*, etc. As we have seen, however, we could equally well turn the problem around, and treat retrodictive relations. Thus, in a certain fraction of cases, *D* came from *C*, *C* from *B*, etc. But what is even more important is that there could exist over-all features of the pattern of events, which could not be expressed solely in terms of the order of succession of events or of its inverse. To see what could be meant by such features, consider, for example, a distribution of objects on a line in space. If the distribution has a certain kind of simple repetitive pattern, then one could predict the objects to the right of a given point if one knew what was to the left and vice versa. However, more generally it is evident that there could exist quite definite and regular relationships which would not allow such a prediction. We have a similar experience with order in time in a musical composition, in which the whole pattern of notes may be quite definite, while, nevertheless, a given note is not determined completely by those which came before it. What we are proposing is that the fundamental laws of physics may have such features too. And if this should be the case, then they would imply the possibility of new kinds of facts; namely, relationships that cannot be expressed in terms of serial order alone. Facts of this kind must escape our conceptual net as long as we think only in terms of predictions of future events on the basis of initial conditions.

Naturally, if such new forms of relationship in time exist, they will require new theoretical schemes to express them. The whole idea of writing a wave function at a given time and calculating a future wave function with a Hamiltonian would have to be given up (except as an

approximation valid in the limiting case of atomic orders of dimensions, for which the present theory is evidently essentially correct). One would have to come to different conceptions concerning order and relationship in space and time, which evidently cannot be discussed in detail in an article of this nature. However, it can be said that such changes in concept are quite consistent with the general character of the quantum theory and would indeed lead to a more natural expression of them than is now possible.¹

We see then that it may well be important to get out of the habit of thinking that in the expression of the fundamental laws of physics, all facts must take the form of initial conditions referring to a given moment of time, while all laws must take the form of relationships, permitting the prediction of events at a later moment of time.

We now come to the second general habit of thought that was to be discussed here; namely, specialisation. Of course, it is evident that specialisation has been forced on us by the overwhelmingly large mass of complex data available today. It is generally realised that while specialisation has important advantages, it requires us to pay a price to get them. There is, however, a certain aspect of this price, that does not seem to have been considered adequately. This is that specialisation narrows our conceptual framework sufficiently so that we cannot even express facts of a certain kind; namely, those whose framework essentially and intrinsically cuts across the boundaries between various fields. (Here, one is referring to facts of a kind that cannot occur in 'border-line' sciences, which are essentially fields of specialisation that overlap two or more other fields.) If such facts exist and are important, then it may well be that the price of specialisation will eventually involve the erection of a barrier to the very knowledge that we expect to get through it.

To bring out the point in more detail, let us consider, for example, the relationship between physics, chemistry, and biology. Of course, each of these sciences studies its problems in its own specialised way, and it is generally admitted that this is necessary, at least to start with. But by now, there is a growing realisation that these subjects are closely related. The most common form that has been suggested for this relationship is that physics is the 'fundamental' science, which could, in principle, contain the other two, if only we knew its laws deeply enough and could calculate their consequences with sufficient detail.

¹ The author expects to publish some preliminary results of a study of this question in the near future.

In other words, it is supposed that all the facts that the chemist and the biologist deal with could, in principle, be expressed in terms of the concepts of physics (even if this might prove to be so impracticable that it would never really be done in full detail), while none of the facts of physics would ever require the concepts of chemistry or biology for their expression.

If we reflect on this point of view for a while, we will see that it is very similar to several others that we have discussed here. Nineteenth-century physicists thought that all facts could be expressed in terms of Newtonian concepts. Before the hypothesis of non-conservation of parity was considered, it was thought that all facts could be expressed in terms of the concept of the symmetry of space. Even now, physicists generally proceed as if all facts can be expressed in terms of some kinds of initial conditions (although a tendency toward developing other points of view has begun).

The notion that everything is, in principle, reducible to physics has in common with their various points of view the character of being an unproved assumption, which is capable of limiting our thinking in such a way that we are blinded to the possibility of whole new classes of fact and law.

In order to make this point in a still more definite way, let us consider the study of living systems in biology. Now living systems are characterised by a great complexity of organisation and a high degree of integration. It may well be that atomic particles out of which such an organised and integrated system is constituted follow laws slightly different from those obtained by extrapolating the laws applying in systems consisting of only a few particles. These differences might be so small that it would be extremely difficult to detect them except in their over-all cumulative effects in a living organism.

If such were the character of life, then it is clear that the hope of expressing all the laws of nature *solely* in terms of physical concepts could not be fulfilled. In other words, it is possible not only that the concepts arising in the physicist's field of specialisation are not sufficient for a complete treatment of all aspects of Nature, but also, that these concepts may not even be completely adequate for relating all the facts that might arise in the physicist's own field.

The above is not a purely abstract possibility. Indeed, there are specific reasons why it may actually be relevant. For example, if we consider the modern quantum theory, we note that it does not deduce classical theory as a limit, in the same way that relativity theory leads to

non-relativistic theory, when the velocity is small compared with that of light. Rather, quantum theory *presupposes* the applicability of classical theory.

At first sight one might think that the above statement is contrary to the results of certain approximations techniques (e.g. the Wentzel-Kramers-Brillouin), which can be applied, for example, to Schrödinger's equation, to show that with heavy objects, the spread of a wave packet is so small that it can, in general, be neglected for practical purposes. Thus, one would expect to be able to represent a 'classical' object by a wave packet of negligible width, which does not spread appreciably in any period of time that is likely to be of interest.

It must be remembered, however, that there is an important difference between an approximation which is good enough for a certain limited range of practical purposes, and a correct *logical* formulation of the theory. It is well known that as conditions and context change, even what are initially small errors can lead to qualitatively new conclusions. And indeed, there are many conditions and contexts in which the representation of classical objects by wave packets can be shown to be inadmissible, especially those in which it is significant that the wave function generally is multidimensional. An important case of this kind arises in the quantum mechanical description of the process of measurement itself. The observing apparatus is a large-scale and essentially classical system. We consider an experiment in which this apparatus is used to measure some property of a micro-system to a quantum-mechanical level of accuracy. Then, after the process of interaction between the apparatus and observed system is finished, but before any observer has looked at the apparatus to see what the result is, the wave function for the combined system takes the form of a linear combination of packets, covering all the possible results of the measurement. On the other hand, we know that in reality, the combined system is in some *one* of its possible states, and that the observer's act of looking makes no significant change in the state of this apparatus (just because the latter is so heavy as to be essentially classical). Thus, here is a case in which the actual state of a large-scale system cannot be represented by means of a single narrow wave packet, because even the wave function of the large-scale system can cover a significantly ambiguous range of possibilities, when it has interacted with a micro-system in such a way as to make a measurement of the properties of the latter possible. (A similar result could follow for a wide range of

interactions with a micro-system which were not measurements at all.) It is clear that classical systems have certain properties that cannot always be represented by narrow wave packets which do not spread appreciably.¹

Roughly speaking, it may be said that the quantum theoretical concepts imply a certain range of potentialities (described by the wave function) as well as a tendency for each system to cover all of its potentialities. On the other hand, classical concepts express the possibility of a system to settle down to one actual mode of behaviour; that is, to choose one actuality out of all these potentialities. In this way, classical concepts make an essential contribution to the expression of the total content of the theory, which is not reducible to an approximation to what appears on the quantum-mechanical side alone. The correspondence principle then merely asserts the compatibility of the quantum theory with the classical theory in the large-scale limit.

Now, it is clear from the above that the classical theory by itself can be accepted as a good approximation in large heavy systems containing many particles.² Suppose, however, that we consider an intermediate domain, in which systems contain a fairly large number of particles, not large enough to be fully classical in their behaviour, and yet large enough to *begin* to exhibit the classical tendency to settle down to a definite mode of actualising its potentialities. In the intermediate domain, there is room for a new kind of law, which in some way synthesises the classical and quantum-mechanical sides. Such a law could not have shown up as yet in our previous studies, in which careful and precise investigations have really been made only of the laws of systems consisting of a few particles.

Now, it is interesting that this intermediate domain is just about that of the order of size of virus particles, which are on the borderline of life. It is well known that these are essentially large molecules, having some quantum mechanical properties, and yet coming close to the classical limit. A law combining classical and quantum properties would express something analogous to one of the peculiar and characteristic properties of living systems; namely, to have a tremendous range of

¹ This whole question is discussed in detail in D. Bohm, *Quantum Theory*, New York, 1951, Chaps. 22 and 23.

² Provided, of course, that these are not interacting to a significant extent with essentially quantum mechanical systems, in the way described previously in connection with a measurement process.

potentialities and to be continually moving toward various actualisations of these potentialities, without thereby falling into a state in which they have no further potentialities.

If such laws exist, then there could be facts that do not fit into the physicist's frame of concepts. They might well be expressible *fully* only with the aid of new concepts which are, in effect, a definition of some of the essential characteristics of living systems.

The above discussion shows in some detail that there is a distinct possibility that the current tendency toward specialisation may defeat its own purpose. It is true that by specialisation we can probably accumulate more and more knowledge of a certain kind; namely, details appropriate to the various fields. But by doing this, we may be leading ourselves away from another category of knowledge, a category that may be far more interesting and more important for our general welfare.

To sum up this talk, we wish to call attention to the relationship between the methods of scientific research, and the content of scientific knowledge. The method must be tailored to the content; and if one loses sight of this, one is in danger of being artificially limited in a way that easily escapes conscious realisation. Method is determined in part by the effort to ask relevant questions in our researches; and it is essential to understand that the relevance of a question depends on the character of the material under investigation. Such questions help determine the forms of the facts that can be elicited in further researches. These questions are, in general, limited firstly by our concepts, laws, and hypotheses, and secondly, in a less obvious but equally important way, by our general habits of thought. Such habits can easily blind us to the need for altering our ways of thinking in accordance with the nature of the material under investigation as we penetrate into new domains.

In order to avoid the dangers that have been discussed here, it would seem that it is desirable to re-establish what was once called a 'natural philosophy'. Of course, it is clear that such an attempt would run against the general current as it is today in science. Nevertheless, it will, in all probability, eventually be necessary, if the full potentialities of the scientific method are to be realized.

H. H. Wills Physics Laboratory,
University of Bristol

COMMUNICATIONAL EPISTEMOLOGY (III) *

MAGOROH MARUYAMA

3 *Limitations of Interpersonal Understanding*

WE consider the limitations of interpersonal understanding under three headings: logical limitations, information-theoretical limitations, and resonances.

3.1 *Logical limitations.* The completeness and closedness of a theory (or interpretation, thinking pattern, etc.) give a natural boundary to the theory. Completeness and closedness are defined as follows:—

A theory T is complete if and only if every sentence constructed with the symbolism of T according to the rules for handling it or its negation is in T.

A theory T is closed if and only if the results of all operations given in T on all elements of T are in T.

Here a 'theory' need not be a deductive theory. And operations are regarded as being inherent in the theory. Instead of stating: 'The set is closed under the operation P . . .' we state: 'The set with the operation P is closed. . . .' The reason is that a culture includes reasoning patterns rather than reasoning patterns operating from outside on cultural elements.

In a complete theory any question constructed by grammatical combinations of words found in the theory can be answered by yes or no except the questions starting with what, how, who, etc. But questions such as 'Who did this?' can be replaced by a set of questions 'Did A do this?', 'Did B do this?', etc., as long as the range of 'who' is included in the theory. As long as no new word or concept is introduced, the theory does not have to expand to include new elements in order to be complete. A complete interpretation is self-sufficient, so to speak. And when one is inside a complete interpretation, its self-sufficiency makes one think that it is the only and correct interpretation. But completeness of an interpretation does not of course guarantee its correctness.

A closed system, on the other hand, does not generate anything new unless new elements or new operations are added. For example, the

* Parts I and II appeared in the preceding two numbers.

set of positive integers together with the operation of adding and multiplying is a closed system because the result of adding or multiplying any two positive integers is again a positive integer. But if a new operation 'subtraction' or a new number -1 (or any other negative number) is given, the whole series of negative integers and zero is generated. (Subtraction can be generated by negative integers and addition.) Then the system of positive integers, negative integers and zero together with addition, multiplication, and subtraction is again closed. It cannot be enlarged again until a new element such as a fractional number or an imaginary integer, or a new operation such as division or extraction of square roots is introduced.

The system of communicable reality is likewise closed. Suppose someone believes that the universe is the universe of communicable realities. What he does in his activities is to start with communicable entities and then manipulate them and combine them, using the operations which are in his universe, i.e. those operations and their results which are definable in communicable terms. The products of his activities are again communicable. In his system he never encounters anything incommunicable, nor does he need anything incommunicable. He lets in no contradiction by rejecting everything incommunicable.¹

Western rational thinking, based on the universe of communicable entities, lets in no contradiction by not considering incommunicable or irrational realities. But it would get contradictions if it took them into consideration. It is a closed system. In order to enlarge this closed system, we have to introduce something from outside the system. In fact, we need not go to non-Western cultures to find a new element. In Western cultures, psycho-analysis and psychiatry as well as poetry have found the existence of incommunicable realities such as the unconscious. Philosophy may keep a respectful distance from them. But their existence makes it no longer possible to deny the fact that the rational system is closed. This awareness of the rational system (or the system which is called 'rational' in Western cultures) as a closed system may logically eliminate some of the prejudices against non-Aristotelian logics as 'illogical'.

A question may be raised here as to the possibility of full communication between two different thinking patterns starting from the intersection (i.e. the common parts) of the thinking patterns, provided that the intersection is not empty. The answer is that, if the two thinking patterns are closed systems, their intersection is also closed,

¹ See Part I, first reference

COMMUNICATIONAL EPISTEMOLOGY (III)

and therefore by starting from nothing but the elements and operations given in the intersection, nothing outside the intersection is gained, and consequently the parts which are outside the intersection remain incommunicable to each other. Let A and B be the two closed systems and D their non-empty intersection. Then the elements and operations in D satisfy two conditions: (a) they are in A, and (b) they are in B. By virtue of (a) whatever they can generate is again in A; and, by virtue of (b), whatever they generate is again in B. Therefore, whatever they can generate is again in D.

When two persons, one having a closed and the other an open thinking system, try to start from their common ground, the chances are that the person with the open system will go beyond the common ground and get into the other person's thinking outside the common ground, while the person with the closed system will remain within his system. The intersection is half-open, i.e. it is open from the open to the closed side but closed from the closed to the open side. For example: Let a person A have a closed system of the universe of communicable entities, and another person B an open system of some incommunicable entities and some communicable entities. (We can think of A as a Western philosopher and of B as a Chinese poet, for example.) As A's thinking is based on communicable entities, with enough time and effort A may succeed in educating B and making B learn A's thinking by communicating with him. But even after B has learned A's thinking, the incommunicable part of B remains incommunicable to A. (As we shall see later, incommunicable realities can be 'resonated' between individuals having the same realities. But since A has no incommunicable reality he cannot resonate with B's incommunicable reality.)

Thus a person with a closed thought system cannot understand anything beyond his system by manipulating the elements and the operations given in his system. Only by accepting elements or operations from outside his system can he go beyond his closed boundary. On the other hand, a person with an open thought system may generate new elements and operations by manipulating the elements and operations within his thinking system. But this does not mean that he can generate every element and every operation found in other systems. As we have seen, the system of all positive integers and one negative integer, together with addition and multiplication, is open. It can generate all negative integers, zero, and subtraction. But after having generated them it becomes a closed system; it does not generate

fractions, imaginary numbers, division or the extraction of square roots.

Thus, an open thought system cannot necessarily expand itself to embrace a different thought system. The surest way to expansion is to introduce the elements and the operations found in the second system to the first system, even if the first system is open. Willingness to accept unfamiliar elements and operations which cannot be accounted for in one's own thought system is necessary for the growth of any thought system.

Another hint we get from mathematics is the concept of analytic continuity in the theory of functions of complex variables. It concerns the representation of a function by a power series. The representation by a power series (Taylor's expansion) of an analytic function in complex variables cannot go beyond the domain of convergence of the series. By choosing within the domain near the boundary of convergence another point of expansion and then constructing another power series one may continue the representation of the function beyond the domain of convergence of the first power series. This process may be continued to fill a connected region within which the function is analytic.

In other words, a power series covers only part of the region of the function. But by choosing another centre for another power series near the boundary of the convergence of the first power series but within the boundary, one may cover a part beyond that boundary. The two power series represent the function in different parts with some overlapping, and they coincide where their domains overlap. By repeating the process we can cover a whole connected region within which the function is analytic.

Similarly, it may be possible to cover a region of thought by a set of interpretations, each of which is not enough to cover the whole region, and each of which originates from another to form a lattice.

A question which may arise from our discussion of the logical limitations of interpersonal understanding is: 'As human nature is everywhere the same, is not anyone able to understand any other person's thinking pattern?' Our reply to this question follows.

A large part of our 'thinking pattern' is formed in infancy in the pre-logical and pre-scientific stage and is later reinforced by our childhood experience in the culture in which we grow up. And so to think like a Balinese, for example, we have to divest ourselves of our rational, 'logical' mode of thinking, return to our pre-scientific and prelogical mode of thinking, and begin again along the Balinese

line of personality development. But the process of personality development is largely irreversible. Though we are all born with the same human nature, we are reared in different cultures and become different; the differences are never completely overcome by 'communication'. We may describe the behaviour of the Balinese with dimension reduction and projection, but this is not 'understanding' the Balinese pattern of thinking. We cannot attain *tabula rasa*. An electronic brain can attain *tabula rasa* by erasing all its stored information, but it has a specific logic built into it and can never attain the pre-logical stage.

3.2 *Information-theoretical limitation.* Here a knowledge of the basic ideas of information theory is assumed. The internal state of the transmitter T is encoded into a 'message' which consists of verbal and nonverbal, conscious and unconscious behaviour of T. When this message reaches the receiver R, and before it is decoded by R, there is some loss of information and some distortion. The loss of information is caused by such factors as language difficulty, unfamiliarity of R with T's behavioural pattern, aspects of the message which escape R's attention because of cultural or personality differences between R and T, etc.; all these losses fall under the category of 'noise'. On the other hand, such distortions as social perception, projection, dimension reduction, may be considered either as noise or as a part of decoding. It is tempting to define non-singular (reversible) distortion as decoding, and singular distortion as noise. But the distinction cannot be applied in our case because decoding functions and their inverse functions do not have to be univalent functions.

Let x be T's internal state, f be T's coding function, w be $f(x)$ or T's behaviour, z be the part of w which reaches R after interference by noise, g^{-1} be R's decoding function, and y be $g^{-1}(z)$ or R's interpretation of z . Then the quantity K, where

$$\begin{aligned} K &= H(w) - H_z(w) = H(z) - H(w|z) \\ &= H(z) + H(w) - H(z, w) \end{aligned}$$

H designating entropy,

is meaningful as a rate of communication only if f , g and their inverse functions are univalent, giving $H(x) = H(w)$, $H(y) = H(z)$, $H(x, y) = H(w, z)$, and only if f is known to R and g is known to T. When these conditions are satisfied, w and z in the equation can be replaced by x and y respectively. But as these conditions are usually not satisfied, the quantity K is meaningless. Nevertheless we can say that equivocation results from 'noise' and from an excessive rate of

communication, i.e. if there is noise in the communication system, or if the rate of information transmitted exceeds the capacity of the receiver, then equivocation occurs in R. Thus the capacity of the receiver sets a limit to the rate of interpersonal communication.

Another important limiting factor is the background information. In interpersonal interaction there is usually a body of common background information in the light of which a new message makes full sense. For example, if in the American culture someone distributes cigars in his office, one knows, because of knowledge about the custom in the culture, that a baby has been born to that person, that he must be happy and busy, that his wife needs help, that he may not sleep well, and so on. Or such background meta-information as 'A tends to make pessimistic statements', or 'B does not usually make insinuations' helps one to evaluate the message adequately.

Mathematically, the gain of information from a message is proportional (or is additive on a logarithmic scale) to the relevant background information if there is no redundancy between the message and the background information. In other words, if no part of the message overlaps with what one can infer from one's background information then doubling the amount of relevant background information doubles the amount of information one gets out of the same message. The informative gain of the message is not determined by the structure of the information bound in the message alone, but also depends on the background information of the receiver.

For example, a professor of chemistry and one of his students may both read a report on a new chemical substance which they have not known of before. Both get the same information from the report. (In chemistry there is no such thing as a nuance of interpretation as there may be in reading poetry.) But the professor, having more relevant background information than the student, can extract from the same report some additional information; he may the more fully appreciate the ingenuity of the methods used in preparing the substance, the applicability of the method to other preparations, the competence of the reporting chemist, and so on. The student is not yet capable of extracting all this information from the same report. On the other hand, if there is redundancy between the report and the background information of the professor, i.e. if the professor could infer from his background information a part or whole of the report, the information value of the report to the professor decreases because of the redundancy. The student, whose little background information

has no redundancy with the report, may gain more information from the report than the professor.

The importance of background information becomes apparent when a visitor arrives in a foreign culture in which he has no background information about the conventions, the speech, and behavioural patterns of the people. He bitterly realises that everyone is an unwitting expert in the way of life of his own culture. And his unconscious and unwitting assumption of the pattern of thinking in his own culture, so basic to him that he never notices its existence, can be more harmful to him than ignorance. In such a case, the 'wrong' background information functions as noise or interference in communication. Thus, the absence of the necessary background information and the presence of 'wrong' background information limit interpersonal understanding.

3.3 *Resonance*. Here it is proposed to consider, very briefly, only five distinguishable types of resonance, though there may well be others.

Intellectual resonance is resonance dependent upon capacity of perception, of decoding and internal process as well as the background information of the receiver. When the message exceeds the capacity of the receiver, equivocation and loss of information result. As we have seen, the information gained from a message depends on the amount of the background information and its relevance and redundancy with respect to the message. When the communication is properly matched to the capacity of the receiver, i.e. the rate of information input is neither too great nor too small, he is said to be in optimal intellectual resonance with respect to the communication.

When two persons derive the same information from a message (either between each other or from another source), they are said to be in intellectual mutual resonance with respect to the information bound in the message.

The distinguishing characteristic of intellectual resonance is that it depends only on the capacity of the receiver (including his perception, decoding process, background knowledge and internal process). The capacity of the receiver, especially his background information, can be improved by intellectual process. On the other hand, the transmitter can reduce his rate of transmission. Thus between two intelligent persons intellectual mutual resonance can be achieved, at least theoretically, with time and effort. (They do not have to agree with each other's opinions; each has only to understand what the other means.)

Meta-intellectual resonance is best explained by an example. The

Balinese language, spoken in the Bali culture of Indonesia, conspicuously lacks abstract general words. Each event needs a new vocabulary. When a canoe is to be built, the builders first work out the names for 200 parts of the canoe. When the canoe has been completed the builders forget the vocabulary. At the next canoe construction the builders have to work out another set of 200 words which has no relation to the first set of 200 words. Each event requires a new vocabulary, and the vocabulary for an old event has to be forgotten as fast as possible so that the memory can be used again for a new vocabulary. There are words for 'to cut in even slices across the loaf', 'to cut irregularly', 'to cut in asymmetric pieces sideways', and about thirty different ways of cutting a loaf of bread. They are distinguished strictly, and if the transmitter uses a close but wrong word, the receiver is not able to guess what is meant.¹

On the other hand, the Manus language, spoken in a culture in the Admiralty Islands in the South Pacific, is a highly analytical language. There are no figures of speech, no gender, only third-person pronouns, and very few adjectives. It is a cold, accurate, precise language of people who can count to 400,000 without any form of notation. In short, it is close to the language of mathematical logic.

Now we may ask: Can a Balinese communicate with a Manus without difficulty? The answer is obviously negative. The difference between the thinking pattern of the Balinese and that of the Manus is not a linguistic difference which can be bridged by translation. The difference lies in the level of abstract thinking. It is a meta-intellectual difference. If two persons are on the same meta-intellectual level, we say they are in meta-intellectual resonance. Differences in meta-intellectual level limit interpersonal communication.

Modal resonance. Modal resonance occurs between people using the same mode of communication—denotative, connotative, situation-contextual, verbal, kinesic, somatic, etc.

In some cultures, such as the Swedish, the denotative mode of communication prevails. Expression (encoding) and interpretation (decoding) are straight-forward. In some other cultures, such as the Danish, connotative communication, with indirect expression and insinuation, prevails. The encoding may have a set pattern which is known to all those in the culture. In such a case decoding is accomplished without misunderstanding. The complicated rules of encoding

¹ *Transactions of the 7th Conference on Cybernetics*, Josiah Macy Jr. Foundation, 1950 pp. 174-175

COMMUNICATIONAL EPISTEMOLOGY (III)

may become internalised, i.e. inseparable from the thinking process to such an extent that the processes of encoding and decoding become automatic, unconscious; those using it believe that it is the most natural, direct mode of communication and are not able to communicate in other modes. If a person in a culture with this mode of communication receives a message from a person from a culture with denotative communication, the receiver automatically and unwittingly decodes the message according to his own rules of decoding and finds meanings which were not intended. Similarly two persons with different types of connotative communication will read wrong meanings in each other's messages. On the other hand, the meaning a connotative transmitter puts in his message is not properly decoded by a denotative receiver. In all these cases, people may accuse each other of being unintelligent, illogical, unclear, etc. The incorrect decoding due to the differences in the types of connotativity may be considered as a variation of projection. But it may be differentiated from projection; projection is the addition of extra dimensions in the interpretation, and the dimensions added are usually those with which one is preoccupied and are therefore rather psychological. But incorrect decoding which is due to differences in types of connotativity adds no extra dimension to the interpretation. It is analogous to using a wrong key in cryptography. It is culturally institutionalised while projection is individual. But the borderline between the two cannot be drawn sharply, especially because psychological preoccupations may become culturally institutionalised as a mode of communication.¹

In contrast with the connotative communication in the Danish culture, which is unambiguously decodable among the Danes, some modes of communication in the Japanese culture are not decodable even among the Japanese. Different internal states may be coded into the same form of behaviour, e.g. into a smile. It is concealment if not deception. (Westerners sometimes report that a Japanese would tell his company what he believes would please them, that he does not tell them his true opinion. But this interpretation is incorrect. A Japanese identifies himself with the social institution or situation in which he finds himself. When he is with a superior or a guest, he

¹ M. Maruyama, 'The Multilateral Mutual Simultaneous Causal Relationships between the Modes of Communication, Sociometric Pattern and the Intellectual Orientation in the Danish Culture', Mimeographed, 1959; 'The Dane's Concept of Superficiality, Childishness and Generalization', Mimeographed, 1958; 'The Danish Language', Mimeographed, 1957; 'A Test in Intercultural Communication', Mimeographed, 1959; 'The Images of the Swedes', Mimeographed, 1957

identifies himself with the other's opinion, and even temporarily believes it. The Japanese has no continuous individual identity, but lives as a 'social identity' which may change from situation to situation.¹ This does not imply that there is no deceptive communication in Japan. There is much deceptive concealment of shortcomings to 'keep face'.)

A good example of situation-contextual communication is found in the Danish culture. In this mode of communication, the encoding and decoding do not depend on the structure of the message but rather on the situational context within which the message is sent. For example, a Dane may say 'I like it' and mean either 'I like it' or 'I do not like it' depending on the situational context. In order to decode the remark correctly one has to know the situational context including his taste and attitudes. But if one knows the situational context well enough to decode his remark properly, the meaning of the remark is known without the reception of the remark. In this case, verbal or non-verbal expressions cease to be a primary means of communication. People communicate through situational contexts. It may also be remarked that Danes are more interested in reading the state of mind of the sender of the message than in the contents of the message.

Again, in some cases, apparent communications do not have communicative purposes. An insult may not be intended to insult the receiver of the message: it may serve merely as an outlet for the sender's frustration. In the Danish culture, when the person insulted does not understand the insult (which is usually expressed indirectly) the insulter is nevertheless satisfied with the feeling that the insulted person is not perspicacious enough to understand the insult and is therefore inferior to the insulter. The satisfaction of the insulter is still greater if there is a third person present who understands the insult while the person for whom the insult is intended does not. In this way insults can be made which would imperil the insulter if the insulted understood the insult. Many of the practical jokes played by the Danes on German soldiers during the German occupation of Denmark were of this kind. The Germans could not make 'head or tail' of the jokes; they did not even realise that they were being insulted. The Danes were delighted.

¹ M. Maruyama, 'Air-sucking as Regressive Oral Gratification in the Japanese Culture', *Review and Newsletter*, Section of Transcultural Psychiatry, McGill University, Montreal, 6, 12-13

COMMUNICATIONAL EPISTEMOLOGY (III)

In some other cultures, messages are deceptive and are expected to be deceptive. In such a culture those who conform with the conventions of deceptive communication are not deceptive. On the other hand, a foreigner with a differently deceptive or even non-deceptive mode of communication is deceptive because he does not conform with the conventions of communication within the culture. And when a person in such a culture gives a foreigner a deceptive message expecting him to figure it out, or at least to be prepared for a deceptive message, and the foreigner does not even suspect that the message is deceptive, then the foreigner misunderstands the sender of the message as well as his expectation, and the sender deceives the foreigner unwittingly. Thus, persons with differing deceptivities (including absence of deceptivity) misunderstand each other and accuse each other of being deceptive and dishonest.¹

In other instances messages have no communicative but only a vague suggestive function. In Japanese poetry every preciseness is avoided. A vague theme and an atmosphere may be suggested in very few words, and nothing more. It is the reader's task to interpret it. The more ambiguous the suggestion and the more numerous the possible interpretations the better the poem. The ego is all-pervasive and dissolves in the atmosphere. But in Western poetry a precision of feelings and emotion is evoked. The ego acts towards, opposes, distinguishes itself from its environment and other individuals.

A normal adult has learned to use verbal communication, kinesic communication (through behaviour), praxic communication (through action and performance), and somatic communication (through physiological manifestations). These modes of communication develop at different stages of a person's development. If, at the time of its development, a mode of communication attracts no satisfactory acknowledgment from others or gives no gratification to the person exercising it, the person will later tend to avoid this mode. If two persons rely chiefly on different modes of communication they will find mutual understanding difficult to attain.²

Experiential resonance. Some realities can be understood only by living through them. Among them are certain states of happiness, certain degrees of seriousness, certain senses of humour, aspirations for freedom, the spirit of the Hungarian revolution, and suicidal despair.

¹ P. Linebarger, *Psychological Warfare*, 1955. Also E. W. Barrett, *Truth is Our Weapon*, 1953; also Part I, second reference

² J. Ruesch, *Disturbed Communication*, 1957, 80

Such realities cannot be communicated directly, but can be resonated, or triggered by some cues, between persons who have experienced them. These experiences are objective in the sense that they are real to those who experience them and that they can be shared by many persons. They are not objective in the sense that they cannot be communicated to all persons, and that they cannot be described scientifically.

Relevance resonance. Communication fails if a message, considered important by the sender, is discarded as irrelevant by the receiver. Relevance resonance is similar to intellectual resonance in that it requires some common intellectual background between the communicating persons. It differs in that it does not require exact matching of intellectual capacity and that it requires similarity of interests, goals, and points of view. Further, relevance resonance is related to experiential resonance; if there is experiential resonance between two persons, it is likely that this will result in relevance resonance (but not vice versa). Two persons who have suffered from the same kind of difficulty, extreme hunger or suicidal despair, or who have participated in some intensely emotional experience, have an experiential resonance between them; this sameness of experience may well lead to a sameness of interest, goal, and point of view and thus result in relevance resonance. This example may suggest that experiential resonance is related mainly to the past and relevance resonance to the future. But experiential resonance can be related to anticipation and relevance resonance to past experience.

4 *Concluding Remarks*

We have now completed an introductory survey of the three stages of our inquiry into communicational epistemology. For illustrative purposes we constructed simple models and discussed them formally rather than empirically. The formalisation served to reduce the treatment of the problem to precise logical operations and to emphasise that communicational epistemology can be treated in terms of algebra, operators, and functions, i.e. can be encoded into logical and mathematical terms in order to communicate with the philosophers, logicians and mathematicians who think mainly in these terms. But we have to remind ourselves that an epistemological study has to build not only its contents but also its form on empirical data. Traditionally it has been thought that the form of human knowledge could be

COMMUNICATIONAL EPISTEMOLOGY (III)

determined *a priori* without considering empirical data. We do not deny the existence of some form of human knowledge determinable *a priori*, but we have to realise that much of the form considered in the past to be determinable *a priori* has turned out to be an *a posteriori* internalised process, e.g. the action-state dichotomy, the subject-predicate relationship. Empirical studies are therefore indispensable both for acquiring the content of epistemology and for determining the boundaries of its *a priori* form.

When constructing a formalised model, we should always have the empirical situation in mind. It is the empirical situation which suggests the model. A model is necessarily an oversimplification of the empirical situation. But by having the empirical situation in mind, one knows exactly where and how it is oversimplified, and consequently where and how it can be applied. On the other hand, a model constructed with little or no empirical study remains on the level of oversimplification, and the constructor is unable to determine where and how it can be applied. To arouse interest in the theory we have presented here a discussion with little empirical material to support it. But the direction of the study of communicational epistemology should be from intensive empirical studies to theoretical formulations. Some of the results of empirical studies have been published.¹

The present discussion has been based on the assumption that there are cultures with different thinking patterns, encoding and decoding functions; a summary of some observed differences follows.

Examples of contrasted patterns of thinking, encoding, and decoding functions.

(1) The communication pattern in the Danish culture contrasts sharply with that of most of the Western cultures. Expressions of one's opinions and feelings are considered to be impolite, aggressive, and childish. It is likewise considered aggressive and impolite to ask others about their opinions, feelings, and professional interests. Good conversation consists in manipulating trifling matters of daily life with humour. Introspection, projection, and subjectivism are mature forms of thinking while the accumulation of facts and objective knowledge and the expression of opinions belong to adolescence and are therefore regarded as immature and superficial. Communication is mainly connotative and situation-contextual. The Danish culture provides a sharp contrast with the intensive intellectual

¹ See Part I, second note

intercourse of the French and German cultures, with the accumulation of universal knowledge and articulate expression in France, with thoroughness and systematisation in Germany, and with the denotative mode of communication prevailing in Sweden. In Denmark a respect for isolation and non-interference is a form of considerateness, and this attitude interrupts and reduces communication, especially with reserved new acquaintances and withdrawn individuals.

(2) In the Chinese pattern of thinking there is no concept of substance, and the subject-predicate relationship is dispensable.¹

(3) Hopi metaphysics is based on the distinction between what is manifest, or accessible to the senses, and what is not. The manifest includes experienced events in the past and present with no distinction between past and present. The 'unmanifest' includes future events, secondhand information (past or present) and all that is mental. As an event at a distant place is heard of only after a period of time has elapsed, the more distant the event the more remote in time it is. There is a clear distinction between myth (which is mental) and reality (which is manifest).²

(4) To the Chichewa of East Africa the past recorded in memory is distinguished from the past recorded in the external physical world.³

(5) Cœur d'Alène distinguishes between growth (maturation of an inherent cause), addition from without, and something affected by addition from without.⁴

(6) There is a great difference between the levels of abstraction in the thinking patterns of the Balinese and the Manus.⁵

(7) English understatement appears arrogant to Americans; American overstatement appears boastful to the English.⁶

(8) The Manus find goals and security in activity, the Chinese in stability.⁷

(9) In a culture in which motives cannot readily be put into action, action is more superficial than motives. In a culture in which motive

¹ T. S. Chang, 'A Chinese Philosopher's Theory of Knowledge', *E T C.*, 1952, 9, 203-262

² B. L. Whorf, *Language Thought and Reality*, 1956

³ Ibid.

⁴ Ibid.

⁵ *Transactions of the 7th Conference on Cybernetics*, 1950

⁶ M. Mead, 'An Application of Anthropological Techniques to Cross-National Communication', *Transactions of the New York Academy of Science*, 1947, series 2, 9, 133-52

⁷ M. Mead, *Co-operation and Competition among Primitive Peoples*, 1937; *New Lives for Old*, 1956

COMMUNICATIONAL EPISTEMOLOGY (III)

can readily be put into action, motives without actions are superficial.¹

(10) In a culture in which concealed maliciousness is a real danger, failure to perceive it means wrong social perception. In a culture in which no such maliciousness exists, lack of suspiciousness is not a sign of naivety, as it is in the first culture, but is more realistic.²

(11) In the Arabic culture an emphasis seems to be given to detached details.³

¹ M. Maruyama, 'Critique de quelques idées très répandues au sujet des rapports entre les cultures et la santé mentale', *Revue de psychologie des peuples*, 1959, 14, 273-6

² Ibid.

³ M. Bleuler and R. Bleuler, 'Rorschach's Ink-blot Test and Racial Psychology: Mental Peculiarities of Moroccans', *Character and Personality*, 1935, 4, 97-114; *Memoirs of American Anthropological Association*, 1955, 81

NOTE ON THE AXIOMATIC FORMULATION OF ELECTROSTATICS*¹

RAYMOND A. SANTIROCCO

THE axiomatic foundations of various branches of physical theory have received a good deal of attention in recent years, and it is generally considered of great pedagogical importance to emphasise to the student the conceptual framework within which a particular set of sense phenomena is to be described. Elementary electrostatics is an area where this objective is frequently overlooked. The purpose of this note is to give an axiomatic formulation of the foundations of elementary electrostatics, with emphasis on the empirical aspects of the new primitive signs and relations introduced into the formalism. We also give a proof of the uniqueness of the electrostatic field, making use of an often neglected superposition axiom of the theory.

1 *Introduction*

We shall follow the general lines of logico-empirical analysis in assuming that a theoretical system suitable for comparison with experience must contain three kinds of statements : equations, logical and mathematical rules, and statements relating the symbols in the equations with facts or objects of experience. We will label statements of the first kind either 'axioms' or 'axiomatic definitions', according to the conventional mathematical usage. Statements of the second kind—logical and mathematical rules—will largely be ignored, not because they lack importance or interest but because the chief aim of this paper is to exhibit the empirical content of a physical theory. We shall simply assume that sufficient logical and mathematical rules exist to define unambiguously and consistently all the manipulations we require of the formalism.

Statements of the third kind, relating the primitive elements of the formalism to the results of experience, are essential to any formal

* Received 3.v.59

¹ The author is indebted to Dr E. C. G. Sudarshan for many valuable discussions on the topic of this paper and related topics.

system purporting to be a representation or description of sense phenomena. These statements have taken various forms—the operational definitions of Bridgman and the reduction to protocol sentences of Carnap, to mention just two. The operational definition with its connotation of specific experimental procedure is perhaps closest in spirit to the working method of the scientist, although this point is certainly debatable. In this exposition we shall take the liberty of calling our statements of the third kind ‘operational definitions’, with the following reservations. Bridgman¹ has demonstrated the difficulty of giving explicit operational definitions for common physical concepts under fairly general circumstances; this is not to be gainsaid. It is a curious feature of electrodynamics, however, that most if not all of the elements of the system, although obeying quite different equations from those of mechanics, can nevertheless be given empirical meaning in terms of the operational definitions of purely mechanical² quantities such as force and length. This is in fact the method to be exploited in this paper. We shall presuppose a complete knowledge of mechanics and its associated operational definitions. Our ‘operational definitions’ (although written for the most part in formal language) will take the form of criteria to be met by, and manipulations to be performed with, *numbers resulting from the experimental measurement of mechanical quantities*. For example equation (3), the definition of numerical measure for charge, requires taking the square root of the product of a certain *force* (an experimentally obtained number) and the square of a certain *distance* (another such number). We shall not, however, concern ourselves with the immense practical difficulties of writing down a complete recipe for, say, force measurement. In this implicit fashion the ‘operational definitions’ set up a correspondence between statements of the formal system and the results of experiment. We shall furthermore permit ourselves to use universal quantifiers like ‘for all x ’ in the ‘operational definitions’, with the realisation that the exact procedural meaning of these terms is determined only by the patience of the experimenter.

We may now sketch briefly the types of phenomena to be treated

¹ P. W. Bridgman, *The Nature of Thermodynamics*, Cambridge (Mass.), 1943

² By ‘mechanical’ we mean something normally classed as part of the theory of mechanics. It is remarkable that this relationship between mechanics and electrodynamics can be made to hold in such detail; this fact is perhaps a partial explanation of the extensive effort devoted by Maxwell and others to the construction of mechanical models for electrodynamics.

here. When, for example, amber is rubbed with cat's fur, accelerations of the amber and the pelt are produced which cannot be described by statements couched wholly within the language of Newtonian mechanics. The necessity now arises of constructing a new theory to describe new phenomena, and at this point one customarily starts to talk of 'charged bodies' and 'electrical forces'. Our presupposed knowledge of the theory of mechanics and its associated operational definitions permits us in principle to subtract gravitational, elastic, and inertial forces from the total force observed to act on the cat's fur. Our purpose is to extend the class of forces treated in mechanics to include any force remaining after such a subtraction has been carried out. In the sequel the term 'force' should be understood to mean this residual value. We may summarise by means of a

Language Convention. The non-mechanical interaction between the treated amber and the cat's fur is termed an *electrostatic force*. The bodies so interacting are said to possess *electrostatic charge*. Non-mechanical interaction between the charged amber or cat's fur and other bodies is also said to be due to electrostatic force, and the other bodies are also said to possess electrostatic charge.

We further take from mechanics the concept of point mass, and we shall deal solely with charged point masses in Sections 2, 3, and 4. The empirical meaning of the phrase 'point charge' will be discussed in Section 5 at greater length.

The axiomatic system will then contain in addition to the mechanical concepts 'force', 'length', and 'work', the new primitive sign 'charge' (denoted by q, q', q'', \dots) and the new primitive relation 'equality of two charges' (denoted by ' q equals q' '). The common symbol for equality '=' will be restricted to an arithmetical equality between two numbers. The relation 'equality of two charges' is presumed to possess the usual reflexive, symmetric, and transitive properties. Another primitive sign, 'isolated region', will be introduced to deal with certain problems occurring in charge measurement experiments, and will specify the conditions under which several of the axioms are to hold.

2 Charge Measurement

We introduce the primitive sign 'charge' by means of

Axiom 1. There exists at least one electrostatic charge.

The next axiom of the system is designed to guarantee the possibility

AXIOMATIC FORMULATION OF ELECTROSTATICS

of maintaining a fixed test charge for experiments to be done at our discretion.

Axiom 2. The electrostatic charge on a body kept *in vacuo* remains constant in time, and is independent of the position of the body referred to an arbitrary static coordinate system.

We may therefore choose a unique test charge, denoted by Q , to effect the experiments prescribed in the 'operational definitions' to follow. Our first task is to give such a definition for the statement ' q equals q' '. We are immediately faced with the problems posed by induction phenomena (the distortion of pre-existing charge distributions by the introduction of new charges). We must make sure that the test charge used to probe the interaction of a charge distribution does not modify this distribution so strongly that the measurement loses meaning. The existence of these difficulties is customarily acknowledged by the insertion of such phrases as 'a test charge supposed not to disturb the charge distribution' or 'an infinitesimal test charge' into all definitions involving a test charge. Such phrases can be given very little meaning at this stage in the analysis, since we as yet possess no numerical measure for charge and no criterion for what is or is not infinitesimal. It is preferable therefore to treat induction phenomena by means of an axiom.

Axiom 3. There exists at least one non-vanishing volume element which is an isolated region.

If we are able to empirically define the new primitive sign 'isolated region' as a volume in which induction effects are negligible, we shall then be able to do unambiguous experiments with non-vanishing test charges in such isolated regions. Moreover we will not be required to perform limiting procedures in advance of a definition of numerical measure for charge.

It is convenient to give operational definitions (denoted OD) for 'isolated region' and 'equality of two charges' simultaneously. To make the steps more explicit we introduce first the empirical notion of 'equivalence'.

OD 1a. Two charges q and q' are said to be equivalent if and only if for a fixed $r(Q)$ and for every $r(q) = r(q')$,

$$F(Q, q) = F(Q, q') \quad (1)$$

Here the notation $r(x)$ denotes the radius vector from the origin of a co-ordinate system (in which all charges are at rest) to the charge x , and $F(x, y)$ denotes the force acting on charge x in the presence of

charge γ . The reader will recall that in this and the following OD's, distances and forces are presumed to be well-defined operationally so that unique measurements of these quantities are possible ; a statement involving the sign '=' like ' $r(q) = r(q')$ ' is therefore a statement about the numerical equality of the results of two measurements.

OD 1b. A volume V is said to be an 'isolated region with respect to q and q' ' if and only if for all $r(q)$ and $r(q')$ in V , $F(q, q')$ is an invariant function of $|r(q) r(q')|$.

OD 1c. Two charges q and q' satisfy the relation ' q equals q' ' if and only if they are equivalent in a region isolated with respect to q and Q and also with respect to q' and Q .

The system of OD's 1a, 1b, and 1c defines equality through a process of comparison with a standard; if q and q' separately interact with the standard Q in the same way, they are judged to be equal. The convenience of making charge measurements by means of a standard interaction arises of course from the rôle played by charge in the theory, that of coupling constant or interaction strength parameter.

To convince the reader that the system OD 1 is not unnecessarily complicated we must explore more fully the relation between Axiom 3, with its new primitive sign 'isolated region', and the induction phenomena. The isolated region is plainly a volume element so far removed from free charge distributions, or so symmetrically located with respect to them, that the introduction of charges q and Q into this region does not alter any such distributions. Under these circumstances the interaction between two static charges depends only upon the scalar distance separating them, and is invariant under transformations between static co-ordinate systems. When translated into empirical terms by means of OD 1b, Axiom 3 can be shown to hold in the laboratory. The necessity of introducing the qualification 'with respect to q and q' ' in OD 1b is shown by the following example. Suppose that a charge q is fixed in space, and imagine also a distribution of charged bodies $\rho(r)$ constrained by a mechanical potential U to lie at rest in fixed positions nearby. If we now investigate the interaction of q with another charge q' , we may find that there exist certain surfaces (in some sense symmetrical to $\rho(r)$) along which the condition in OD 1b are fulfilled. The charge distribution $\rho(r)$ does not betray its presence, and as far as this experiment is concerned we have established the existence of an isolated region. Suppose now that we replace q' by a charge q'' so large that $F[q'', \rho(r)] \geq \text{grad } U$. In this case we shall have overcome the constraints on $\rho(r)$ and this distribution will be free

to move and become polarised under the influence of q'' . Then our region will no longer be isolated in the sense of OD 1b. It is therefore necessary to specify for what range of interaction strengths a region of space is essentially isolated.

In connection with the system 1a, 1b, and 1c, several points deserve further mention. First, the notion of equivalence has been introduced only to clarify the steps in the definitions, and possesses no empirical utility. We may dispense with it entirely by substituting OD 1a into OD 1c. Second, it should be reiterated that the OD's 1a, 1b, and 1c contain all the empirical meaning in the statement ' q equals q' '. After a method for measuring charge has been defined, one may show that these definitions imply the equality of the numerical measures of q and q' , but it must be emphasised that this result is a theorem of the axiomatic system. Third, such definitions as OD 1, in which a statement is given meaning before its separate elements have been defined, occur frequently in physics and mathematics.¹ It may be necessary for the student to convince himself that the technique is in fact consistent, since this is not always immediately apparent.

We have now the requisite logical apparatus for the interpretation of a number of experiments. For example, now that we have a definite meaning for the phrase ' q equals q' ', we may demonstrate empirically that

$$F(q, q') = k r^{-3} r \text{ for } q \text{ equals } q' \quad (2)$$

in a region isolated with respect to q and q' . In this equation

$$r = r(q) - r(q').$$

The constant of proportionality k depends uniquely on the particular pair of equal charges q and q' used in the experiment. This relation is of fundamental importance. Indeed, equation (2) is the central datum for electrostatic charge measurement, since at this point in the analysis a choice of the functional dependence of k on q constitutes an empirical definition of numerical measure for q . Furthermore, the choice is arbitrary within wide limits. We shall make the standard definition by choosing $k(q) = q^2$ and writing

OD 2. The absolute numerical magnitude of a charge q is given by

$$q \equiv |k|^{\frac{1}{2}} = [F(q, q) r^2]^{\frac{1}{2}} \quad (3)$$

where $F(q, q)$ is the magnitude of the force between two equal charges

¹ Some examples are found in Mach's definition of mass, Russell's definition of cardinal numbers, and Cantor's theory of transfinite numbers.

q , and r is their separation; the measurement is to be done in a region isolated with respect to the two charges.¹

With a numerical measure for charge in our possession, we can perform as many experiments as we desire to corroborate.

Axiom 4 (Coulomb's Law). In a region isolated with respect to q and q' ,

$$F(q, q') = qq'r^{-3}r \text{ for all } q, q'. \quad (4)$$

It is evident that the form of charge dependence in Coulomb's Law is dictated by our choice of the function $k(q)$, since for the case $q = q'$ equations (2) and (4) must be identical. If for example we were to assume $|k(q)| = e^{2q}$, Coulomb's Law might take the rather unwieldy form $F(q, q') = P(q, q')r^{-3}r \exp(q + q')$, where $P(q, q')$ is a parity factor expressing attractive or repulsive interaction. Conversely if we take Axiom 4 in the form given above, we are led uniquely to OD 2 as the numerical measure for charge. The essential point is this: we may conclude from experiment only that the interaction is bilinear in some function of the charge, and the choice of this function is left to our discretion.² In this sense Coulomb's Law is sometimes loosely spoken of as a 'definition of charge'.

The utility of OD 2 rests on the ability of the experimenter to discover or manufacture two equal charges. This possibility is certainly empirically established; it is equally certain that one cannot prove within the axiomatic framework that there are any equal charges in nature. Axiom 5, the Superposition Principle, relates the scalar addition of charge elements to the size of their interactions, and thus furnishes a recipe for the regulation of the charge on a particular body. The reader will notice that Axiom 5 has been phrased in terms of infinitesimal increments of charge, and may question the legitimacy of introducing such an expression. It should be clear by this time that the calculus obeyed by the sign 'charge' in the axiomatic system is that of arithmetic—in other words 'charge' is a number. The phrase

¹ This definition neglects the sign convention with respect to positive and negative charge; this may be supplied by choosing the sign of the standard test charge Q to be plus, for example, and assigning a plus or minus sign to q according to whether its interaction with Q is repulsive or attractive. There is also implicit in OD 2 an arbitrary choice of units.

² The same considerations hold for the gravitational mass appearing in the Newtonian gravity formula. In this case, however, the principle of equality of gravitational and inertial mass fixes the functional dependence on mass as a bilinear one.

'infinitesimal charge' has therefore the same meaning that 'infinitesimal number' would have in the integrodifferential calculus. The stipulation that the sign 'charge' obeys the calculus of numbers is in fact one of the logical and mathematical rules referred to earlier.

In Section 4 the electrostatic field and electrostatic potential will be defined axiomatically in the forms of limits or integrals, well-defined operations over the field of numbers, and it will be argued that this procedure is quite essential to satisfactory definition of field and potential. What is perhaps more disturbing is that the OD's for field and potential will also be given in the form of 'limits'. In the following section we shall therefore try to describe an empirical process which can be called 'taking the limit' of a set of experimental numbers.

3 'Empirical Functions' and their 'Limits'

Suppose that we make a series of experiments, identical except for the monotonic variation of one of the system parameters affecting the result of the measurement. The results of these measurements can be considered an 'empirical function' of the variable parameter, and this function can be plotted on a graph as a series of points connected by straight line segments. The curve so obtained frequently suggests the graph of some simple mathematical function. For example, if we charge a capacitor and measure the voltage across it as a function of time during its discharge, we would probably not hesitate to call the curve of voltage versus time an exponential. We would equally readily say that the voltage 'approaches zero as a limit' as the time increases. What are some of the criteria used in making this assertion?

It appears that two of the requirements in the mathematical definition of limit cannot be strictly satisfied in the analogous domain of the 'empirical function'; these are the requirements of *smooth* approach and *arbitrarily close* approach to the limiting value of the function. Smooth approach to the limit implies knowledge of the function at *all* points in some neighbourhood near the limit, but since we can measure only at separated points such complete information on the behaviour of the empirical function must necessarily be lacking. We must decide subjectively whether or not the points on our curve are closely enough spaced to justify drawing a smooth curve through them, or even connecting neighbouring points by a straight line. One ultimate boundary on this grid spacing is the accuracy with which the argument of the empirical function can be measured; there is little significance

in choosing a grid spacing of 0.01 seconds if the statistical accuracy of a time measurement is only 0.1 seconds. We must also ignore the possibility of a non-denumerable infinity of singular points in the empirical function, since there is no guarantee that any of the measurements will coincide with, and thus disclose, any such singular points.

Secondly, we must accept a cut-off on how closely the limit can be approached. There is an unavoidable uncertainty in the measured value of the empirical function, due to both the sensitivity of the measuring instrument and the statistical variations encountered among measurements. For example, if the voltmeter in the capacitor experiment were calibrated in thousandths of a volt, we could assert with equal vigour that the 'limit' of the voltage was not zero but 10^{-4} volts. Since this error does not in practice disappear, we can give the limit of an empirical function only within some specified range of uncertainty. In other words the empirically deduced limit of a series of experimentally-obtained numbers is itself an experimentally-obtained number, subject to uncertainties and requiring the statement of these uncertainties for its full interpretation.

If the limit at 'infinity' of an empirical function is required there is yet another subjective judgment to be made, namely the finite point at which the empirical function is assumed to no longer vary detectably from its asymptotic value. Referring again to the experiment with the capacitor, if the voltmeter needle were to remain stationary for a long time compared to the period of noticeable fluctuation, we might then conclude that as far as this experiment is concerned 'infinite time' had been reached.

The foregoing discussion presumes that the quantities to be measured do not occur in indivisible quanta larger than the sensitivity of the apparatus. If this is not true, and either the empirical function or its argument are observed to vary discontinuously within the accuracy of the experiment, there are two choices open. One may accept the quantum size as the measurement cut-off but ignore it otherwise; this is done in classical electrodynamics and a 'macroscopic' theory is obtained. Alternatively one may choose to modify the theory to encompass quantities which are not permitted the luxury of continuous variation. The Millikan oil drop experiment is an example of a measurement which defines the boundaries of the macroscopic theory, but does not otherwise modify it. On the other hand this experiment does of course illustrate what we call discrete, particulate, or 'microscopic' phenomena.

AXIOMATIC FORMULATION OF ELECTROSTATICS

The foregoing is hardly an exhaustive discussion of the problems of describing empirical data by mathematical functions, but does describe some of the more obvious subjective processes used in making any such association. The fact remains that the technique is indispensable in the exact sciences, and is the most powerful method of constructing from empirical data statements which can be directly compared with statements of the theory. We may therefore expect that if we ask a laboratory worker to make a particular set of measurements and determine what 'limit', if any, they approach at some point, he will in most cases be able to give an answer in the form of a number with an attached specification of accuracy.

4 *Electrostatic Fields and Potentials*

Let us make

Axiomatic Definition 1. The electrostatic field E at a point r as defined as

$$E(r) \equiv \lim_{Q \rightarrow 0} F(Q)/Q \quad (5)$$

where $F(Q)$ is the force exerted on a test charge Q located at the point r .

An OD analogous to this definition can be phrased in terms of an 'empirical function'.

OD 3. Obtain a series of measured test charges Q_i , ranging in magnitude from the largest convenient value down to the smallest which can be constructed. Place each charge in turn at the point r , measure the (electrostatic) force $F(Q_i)$, on each charge, and construct an empirical function of Q_i from the series of quotients $F(Q_i)/Q_i$. By some standard graphical or numerical procedure test the convergent behaviour of this function as Q_i is decreased; the value, if any, approached by the empirical function is defined as the electrostatic field at the point r .

Note that the use of a limiting procedure frees us from specifying that the experiment be done in an isolated region. In essence, taking the limit $Q \rightarrow 0$ makes the entire space an isolated region. If restricted solely to finite isolated regions our definition would fail in the cases of greatest interest (regions near charge distributions), and the field concept would lose its tremendous power as a descriptive scheme. The significance of the field concept in classical physics is that the field can be considered to store and to transport energy and momentum, thus preserving the conservation laws in action-at-a-distance phenomena. The ability of the field to maintain the integrity of conservation laws depends on the field's being continuously defined everywhere,

with the exception of possible singularities at the field sources. Furthermore, since a given volume is isolated only with respect to a certain range of interaction strengths, the electrostatic field defined in terms of isolated regions would in general be multivalued. Therefore all restrictions relative to isolated regions must be eliminated in the field definition, and for this purpose the use of a limiting procedure seems unavoidable.

It is not evident *a priori* that the limit in equation (5) exists, and this question is worthy of an explicit answer. Our proof depends on a frequently overlooked linearity axiom of the theory. We now introduce

Axiom 5 (Principle of Differential Superposition). Let $\{\delta q_i\}$ be a set of infinitesimal charges. Then

$$F(Q, \sum_i \delta q_i) = \sum_i F(Q, \delta q_i) \quad (6)$$

where $F(Q, \sum_i \delta q_i)$ represents the total force on Q due to $\{\delta q_i\}$.

We may now prove

Theorem 1 (Uniqueness of Field). The limit described in Axiomatic Definition 1 exists almost everywhere.

From Axiom 5, the interaction of the charge distribution to be described by a field E can be represented by the interactions of some set $\{\delta q_i\}$, the choice of which set does not depend on Q . Furthermore, for all Q less than or equal to some Q_0 the region around the point r is isolated. It then follows from equations (5) and (6) that

$$E(r) = Q_0 \lim_{Q \rightarrow 0} \left[\sum_i F(Q, \delta q_i) \right] / Q_0 \quad (7)$$

and using equation (4) we find

$$E(r) = \sum_i \delta q_i |r - r_i|^{-3} (r - r_i) \quad (8)$$

Thus $E(r)$ exists except at the points $r = r_i$, and the desired theorem is proved.

Our final objective is to introduce the concept of electrostatic potential. Let us make

Axiomatic Definition 2. The electrostatic potential at a point r is defined as

$$\phi(r) \equiv - \int_{\infty}^r E \cdot dl \quad (9)$$

where the line integral is along an arbitrary path.

AXIOMATIC FORMULATION OF ELECTROSTATICS

It is easy to show from Axioms 4 and 5 that curl E vanishes, and from this there follows

Theorem 2. $\phi(\mathbf{r})$ exists and is single-valued.

For purposes of clarity we give the OD for electrostatic potential in two steps; for brevity we shall use mathematical notation and leave to the reader the elaboration of OD's 4 and 5 in language similar to that of OD 3.

OD 4. The electrostatic potential difference between the points \mathbf{r}_1 and \mathbf{r}_2 is

$$\text{PD}(\mathbf{r}_1, \mathbf{r}_2) \equiv \lim_{Q \rightarrow 0} W_Q(\mathbf{r}_1, \mathbf{r}_2)/Q \quad (10)$$

where $W_Q(\mathbf{r}_1, \mathbf{r}_2)$ is the work done on a charge Q in moving it quasi-statically from \mathbf{r}_1 to \mathbf{r}_2 by an arbitrary path.

OD 5. The electrostatic potential at a point \mathbf{r} is

$$\phi(\mathbf{r}) \equiv \lim_{r' \rightarrow \infty} \text{PD}(\mathbf{r}', \mathbf{r}). \quad (11)$$

Note again that the use here of the limiting procedure is essential. OD 4 may be given without a limiting procedure by specifying that the experiment be done in an isolated region. Then, however, OD 5 will fail, since we cannot be sure that the arbitrary path of OD 4 will always pass through isolated regions as $r' \rightarrow \infty$.

5 Point Charges and Coulomb's Law

Our discussion so far has been restricted to interactions among point charges, and Coulomb's Law has been represented as exact only when applied to point charges. In a sense our axiomatic system is an implicit definition of the notion 'point charge'. It is therefore of interest to investigate what empirical meaning this concept may have, apart from being a point mass (a well-defined term in mechanics) bearing an electrical charge. Intuitively we understand by the phrase 'point charge' a charge distribution with vanishing spatial extension; we may then expect that the exact Coulomb's Law is in some way a limit of the interaction laws for extended charge distributions.

Consider a sphere, bearing a charge q , in an isolated region whose dimensions are large compared to the diameter of the sphere. We shall probe the interactions of q with a test charge Q which we assume to have small dimensions compared to the sphere (this restriction is convenient but not essential). Since the region is isolated we know that Q does not noticeably perturb the charge distributions external to the region, but we are not assured that Q leaves the distribution of q

undisturbed. If the sphere is a conductor, the distribution will in fact polarise until the surface of the sphere is an equipotential. Under such circumstances Coulomb's Law does not hold for the interaction between q and Q . Now the charge distribution on the sphere may be expanded in a series of multipoles where the potential of the n th multipole depends as $r^{-(n+1)}$ on the separation r between sphere and test charge. As r is made larger, the effects of the higher order terms in the expansion decrease until at last only the monopole term contributes to the interaction. In this limit the interaction measures only the product of the test charge Q and the sphere's total charge q , and Coulomb's Law holds with arbitrary precision. If we relax the restriction on Q and allow it to be distributed over a body comparable in size to the sphere, it too will become polarised. However, the distribution of Q can also be given as a multipole expansion and the argument carries through as before.

This limiting process rigorously eliminates all shape-dependent effects on the source of the field, and permits isolation of the 'true' electrostatic interaction. At the same time it can be taken as an empirical definition of the term 'point charge interaction'. The question of infinite point-charge self-energies is also to some extent avoided, since the procedure does not permit arbitrarily close approach to the field singularity of the point charge.

We see therefore that the interactions between charge distributions of finite extension obey Coulomb's Law as a limit. Note that this procedure does not apply to infinitely extended charge distributions; one cannot for example define operationally the total charge on an infinite conducting plane. This fact corresponds to the analytical statement that in general the monopole moment of an unbounded charge distribution does not exist.

It may be of interest to make one further general observation. In the case of electrostatics, as in other portions of classical theory, most of the fundamental axioms are amenable to direct experimental verification; we may identify statements of fact corresponding to the axioms by means of the OD's, and examine these pairs for consistency. In the more recent theories of modern physics empirical statements corresponding to the axioms are frequently inaccessible to direct experiment, and even meaningless in some cases. In the special theory of relativity the axiom 'The velocity of light is the same in all inertial frames' cannot be directly tested in many reference frames of great interest (such as the rest frame of a relativistic unstable particle) because

AXIOMATIC FORMULATION OF ELECTROSTATICS

of the great experimental difficulties. In quantum theory the situation is even more remote. Here the elements of the axioms (operators and wave functions) have *by definition* no OD's and therefore no empirical counterparts susceptible to measurement; the theory says that they are not 'observables'. One cannot speak of an experiment corroborating Schrödinger's equation in the same sense as an experiment corroborating Coulomb's Law; one can only do an experiment corroborating some theorem deduced from Schrödinger's equation.¹ In these cases the physicist must be content to glean empirical support for his theories from verification of certain theorems rather than axioms. This of course does not detract from either the utility or significance of these theories, but may perhaps be a standard by which one theory is judged more 'abstract' than another.

Research Division
General Dynamics/Electronics/
Rochester, New York

¹ Of course due to the tautological character of the theorem it can be viewed merely as a restatement of its premises with nothing added to their content. The statements of premises are however more remote from experience (in quantum theory) than many of the theorems, and there is always the possibility of error (or more commonly, inadequate approximations) in the derivations.

NOTE AND COMMENT

'Lying' and the Compleat Robot

MICHAEL SCRIVEN has tried, with great subtlety, to indicate what the complete robot will be.¹ He has even suggested how far one can go in eliciting the answer from the robot himself (or herself)² if it is built in a certain way. But I think the question whether a robot can lie requires some further discussion than the extremely casual one given by him in his study.

'Lying' involves telling something different from what one knows to be the case. This, of course, is only a first approximation. Lying is by no means confined to verbal reporting only. One may lie in a hundred ways. The gesture, the expression, the turn of an eye, the inflexion of a voice all may be used for lying even more effectively than the utterance of a sentence. Rather, a sentence which ostensibly is false may be uttered in such a way as to convey just the opposite. Or, equally, a true sentence may be embedded in a context of expressive gesture which turns it into an effective lie. 'Lying', then, is a rather complex affair and needs exploration both on the intentional and the performatory side before the Compleat Robot may do what he is expected to do.

One need not exactly *know* the case, in order that the lie may be a lie. One's belief for that purpose is sufficient. If I believe that such is the case and try to produce, by whatever means, a belief contrary to that which I believe, this will be regarded as 'lying' even if my own belief is *actually* mistaken. In such an event, it will be difficult to suspect that what I told was a lie, for it will be found to be true and normally taken as having been uttered in that way. It is not, thus, the *actual* contrariety of what is uttered or expressed to whatever happens to be the case that makes a lie a 'lie'. Rather, it is the contrariety to what is *believed* to be the case that turned it into a 'lie'. And 'believing' is a very, very 'intentional' affair. There is, of course, a performatory aspect to 'belief' and we do try to square a man's professed beliefs with his behaviour. We even go so far as to infer unconscious beliefs that seem to be involved in the behavioural choices of persons. But, firstly, one does not generally speak of the beliefs of animals, not to speak of the beliefs of robots. There seems something incongruous in saying that the behaviour of the animal implies that he believes in such and such beliefs. For example, the statement 'my dog does not seem to believe in an after-life' makes no sense as its opposite doesn't do either. This may

¹ *Dimensions of Mind*: A symposium, ed. by Sidney Hook, New York, 1960.

² The question of sex in robots has, as far as I can see, not been discussed up till now.

suggest that the ascription of belief to some object can make sense only if it can be said, in some sense, to disbelieve also. Secondly, there seem to be beliefs whose performatory aspect is too much confined to their verbalised expression. For example, if someone says that he believes in ghosts or that there was a life before birth or that there will be life after death, it is not quite easy to see what further performatory behaviour on his part is *necessarily* required for our believing in his belief. The person believing in ghosts may try table-tipping or show fear in the dark but he may, equally well, not do so. Yet, the mere verbalised expression on his part is, in most cases, sufficient to induce us to believe in the truth of his statement. On the other hand, no one will believe in this belief if a machine was made to utter such a verbalised expression.

The difficulties with regard to 'lying' may increase still further if we take into account the degrees of belief which so usually characterise our psychical life in this respect. I wonder if it will be called a lie if I try to produce a greater sense of conviction in the other person than I myself possess in the matter. This introduces another aspect which may be regarded as intrinsic to lying, that is, it has a purpose or an end to achieve without which it will not be communicated at all. I may be mistaken about the way I adopted to make my lie succeed in its function, but that I do want it to succeed is the very *raison d'être* of lying. A pure lying for its own sake will not be lying. It will be just joking or having fun or imagining or telling a tale or anything else. The point of lying is that the other person should not be able to guess that it is a lie and the success of lying can only be determined if we know what the lie was for.

These considerations, I think, are extremely important for even the performatory aspect of lying to which Michael Scriven has confined himself in his paper. The point is how can one determine the performatory aspect unless, to some extent, one has determined what 'lying' is? Scriven, I am afraid, has not even raised the issue in his paper. He writes: 'We also introduce the robot to the concept of truth and falsity and explain that to lie is to utter a falsehood when the truth is known. . . .'¹ This obviously is an extremely loose kind of statement. One does not introduce robots to 'concepts' nor does one 'explain' anything to them. One just builds into them structures which perform the behavioural responses which we consider to be relevant to the doing of a certain action or being a certain sort of activity. The proper question would then have been to ask what is the performatory activity which we would have to build in a machine so that it may be said to 'lie' when it performs that sort of behaviour? Michael Scriven has not even raised this question, though it is so central to his whole argument. Perhaps he thought the matter quite simple, not worth a detailed discussion at all.

¹ Michael Scriven, op. cit. p. 141

However, if we do think it out, there seem to be difficulties which do not seem quite so easy to overcome. This may, of course, be overcome or circumvented but it has to be shown how. For example, following Mr Scriven's explanation to the robot, we may build in a structure which would make the robot always say the opposite of what it 'knows'. Supposing we do it, would this be 'lying' as we know it in a human being. Surely not, for here the lying would become so mechanical that the person noting the robot's responses would quickly find that the robot is functioning differently and will correct his interpretations and behave accordingly. The more important point, however, is that this will not change the robot's behaviour at all. It will continue to act or speak the lie which has been built into him even when it (the lie) has ceased to perform its function. We may, of course, try to build a robot which will change its behaviour on noticing that its lie does not work. But what will this exactly mean? How will the robot notice, say, my disbelief in his reporting. I need not do anything at all to disbelieve what someone is saying. Also, the range of actions that I may do as a result of my disbelief are too large and varied to be built into the robot's system as stimuli arousing it to change his performatory lying. Not only this, he is to be built in such a way that on getting the stimuli which are the performatory equivalent of disbelieving, he should be able to devise new ways of deceiving the disbelievers once again.

On the other hand, instead of making the lie mechanical we may just 'randomise' it in the robot's system. Now the robot will not lie every time it knows what is the truth, but only on chance occasions. This will hardly be 'lying' at all, for the robot may come to lie at a moment when it interferes with its own interests. In other words, if the robot were conscious in such a situation he will not lie at all. One may, of course, try to exclude such situations by building into the robot mechanisms which will exclude such self-defeating lying on his part. But, then, the rest of the lying would be some sort of a joke or a piece of fictional imagination if it is not also geared to some ends which the robot has been built to achieve.

This, perhaps, is the central point. Can 'lying' be adequately understood even in its performatory aspect without reference to the purpose it seeks to achieve? Michael Scriven has gone on to write: 'We then add a circuit to the robot, at a special ceremony at which we also christen it, which renders lying *impossible* regardless of conflict with other goals it has been told are important.'¹ Professor Scriven has forgotten that if the poor robot cannot but tell the truth, it is a robot indeed. 'Truth-telling', like 'lying', cannot be mechanised, that is, made compulsory without losing all that we connote by these terms in their human context. Further, it is not so easy to tell the truth as Scriven seems to suppose. Even the question what is truth

¹ Michael Scriven, op. cit. p. 141 *italics mine*

'LYING' AND THE COMPLEAT ROBOT

has been the subject of unending debate among those who have tried to think about the matter.

Between the mechanical and the random may lie the area where we may try to build in 'lying' in the service of some purpose which the robot can vary if it finds that it does not succeed in the achievement of that purpose. I wonder what will be the engineering problems set by such a task. But, at least this much seems fairly clear to me that if the robot aspires to be human, it cannot be made impossible for it to tell a lie, though what it will mean for a robot to tell a lie is not clear either. In any case, there seems little doubt that a lot of analysis is needed even on the performatory level to say that a robot is 'lying' or 'telling the truth'. And if Scriven's experiment depends on the impossibility of the robot's telling anything but the truth, his experiment is bound to fail if simultaneously he is trying to make it as human as possible, that is, if it is being made in principle 'unpredictable in a way essentially similar to the way human beings are.'¹ If, on the other hand, the robot's unpredictability is given up in the matter of 'truth-telling' then whatever the robot's answer to Scriven's question, we may rest assured that it is not human in the sense in which we consider ourselves to be human, that is, free to tell the lie or truth as it pleases us.²

DAYA KRISHNA

¹ Michael Scriven, *op. cit.* p. 122

² It is interesting to note the similarity of the robot-maker's dilemma to that of God with respect to man. If man is given freedom, then he may develop all sorts of undesirable qualities and if he is not given freedom, there is no virtue in his love or prayer or truth-telling or anything else that he does.

DISCUSSIONS

MORE ABOUT LORENTZ TRANSFORMATION EQUATIONS

RECENTLY G. H. Keswani has published a paper under this title (this *Journal*, 1960, **II**, 50) in which he tries to establish a contradiction between the principle of Relativity and the Lorentz transformation equations, and infers from this the existence of a preferential system of reference (ether). I wish to point out that these conclusions derive from an elementary error.

He considers (p. 52) two co-ordinate systems $S(x, y, z, t)$ and $S'(x', y', z', t')$ with parallel axes, S' moving in the x -direction relative to S with the velocity v . Then the connecting Lorentz transformation is

$$x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}}, \quad y' = y, \quad z' = z, \quad t' = \frac{t - (v/c^2)x}{\sqrt{1 - v^2/c^2}} \quad (1)$$

Then he considers a spherical light wave spreading from the common origin O at the time $t = t' = 0$ and says (I quote): 'In S , the intercepts of the three axes are

$$x = r, \quad y = r, \quad z = r \text{ at time } t = r/c, \quad (2)$$

He inserts these values in the Lorentz transformation as if they were meant to be the co-ordinates of one point (not of three), and calculates the co-ordinates of this point in the system S' . Thus he finds

$$x' = r \left(\frac{c - v}{c + v} \right)^{\frac{1}{2}}, \quad y' = r, \quad z' = r, \quad t' = \frac{r}{c} \left(\frac{c - v}{c + v} \right)^{\frac{1}{2}}. \quad (3)$$

He concludes that the 'three points' in S' are not on a sphere but on an ellipsoid. What his calculation actually means is the transformation of a single point from S to S' , namely the corner opposite to O of the cube with sides parallel to the axes in S and with the length r , at the time when a light signal from the origin reaches it. There is, of course, not the slightest contradiction against relativity in his results.

To follow his programme, one would have to consider the three axial points at distance r at the same time $t = r/c$, which in S are given by:

$$\begin{cases} P_1: x_1 = r, \quad y_1 = 0, \quad z_1 = 0, \quad t_1 = r/c, \\ P_2: x_2 = 0, \quad y_2 = r, \quad z_2 = 0, \quad t_2 = r/c, \\ P_3: x_3 = 0, \quad y_3 = 0, \quad z_3 = r, \quad t_3 = r/c. \end{cases} \quad (4)$$

These all lie on the spherical wave front

$$x^2 + y^2 + z^2 = r^2, \text{ with } r = ct. \quad (5)$$

The Lorentz transformation leads to

$$\begin{cases} P_1: x'_1 = r \left(\frac{c - v}{c + v} \right)^{\frac{1}{2}}, & y'_1 = 0, \quad z'_1 = 0, \quad t'_1 = \frac{r}{c} \left(\frac{c - v}{c + v} \right)^{\frac{1}{2}}, \\ P_2: x'_2 = -r \frac{v}{\sqrt{c^2 - v^2}}, & y'_2 = r, \quad z'_2 = 0, \quad t'_2 = \frac{r}{c} \frac{c}{\sqrt{c^2 - v^2}}, \\ P_3: x'_3 = -r \frac{v}{\sqrt{c^2 - v^2}}, & y'_3 = 0, \quad z'_3 = r, \quad t'_3 = \frac{r}{c} \frac{c}{\sqrt{c^2 - v^2}}. \end{cases} \quad (6)$$

REPLY TO PROFESSOR BORN

These three events are not simultaneous in S' ; the first P_1 is on the wave-front

$$x'^2 + y'^2 + z'^2 = r'^2, \text{ where } r' = r \left(\frac{c-v}{c+v} \right)^{\frac{1}{2}} = ct', \quad (7)$$

the other two, P_2 and P_3 , on the different wave front

$$x'^2 + y'^2 + z'^2 = \bar{r}'^2, \text{ where } \bar{r}' = r \frac{v}{\sqrt{c^2 - v^2}} = ct'. \quad (8)$$

It is obvious that the radii (and the corresponding transmission times) are not the same for the 'longitudinal' point P_1 and the two transverse points P_2 and P_3 . But this does not, of course, mean that the wave front in S' is ellipsoidal; only its time of expansion is different for the three points when observed in S' .

By drawing a Minkowski diagram in three dimensions x, y, t (omitting z) one can see all this intuitively.

The difficulties met by Mr Keswani in the connection between the principle of relativity and the Lorentz transformations are due only to primitive errors in his calculations. I regret that such faulty considerations are put before the readers of a philosophical periodical as they thus are made to believe that physicists do not know what they are talking about.

MAX BORN

Bad Pyrmont/Germany
Marcardstrasse 4

REPLY TO PROFESSOR BORN

THE question is: how is a spherical wave-front of light in S observed in S' ?

We said that it is a sphere if we consider the principle of relativity, but that it is an ellipsoid if we apply the Lorentz equations for transformation. Born says that the wave-front in S' is not ellipsoidal.

What did the prophet himself say? In his original paper of 1905 Einstein said

The wave under consideration is therefore no less a spherical wave with velocity of propagation c when viewed in the moving system . . . (*The Principle of Relativity, Electrodynamics of Moving Bodies, 1923, p. 46*)

But a little later, when dealing with the transformation of energy of light-rays he stated

The spherical surface—viewed in the moving system—is an ellipsoidal surface . . . (op. cit. p. 57)

These two statements were quoted in our paper but Born does not comment on them.

Einstein obtained an ellipsoid as follows (op. cit. p. 57). Here we use our notation. The equation of the spherical surface expanding with the velocity of light in the 'stationary' system (i.e. S) is:

$$(x - ct)^2 + (y - mct)^2 + (z - nct)^2 = r^2 \quad (1)$$

where l , m , and n are the direction-cosines of the wave-normals. For $t = 0$, we get the space-equation of the wave-front, $x^2 + y^2 + z^2 = r^2$ in S . Putting

$$x = (x' + vt')/\sqrt{1 - v^2/c^2},$$

$t = (t' + vx'/c^2)/\sqrt{1 - v^2/c^2}$, $y = y'$, $z = z'$ and $\beta = 1/\sqrt{1 - v^2/c^2}$, equation (1) is transformed into:

$(\beta x' + \beta vt' - \beta lct' - \beta lx'v/c)^2 + (y' - \beta mct' - \beta mx'v/c)^2 + (z' - \beta nct' - \beta nx'v/c)^2 = r^2$. Setting $t' = 0$, Einstein obtained the following spatial equation of the ellipsoidal surface in the moving system:

$$(\beta x' - \beta l v/c)^2 + (y' - \beta m v/c)^2 + (z' - \beta n v/c)^2 = r^2 \quad (2)$$

Gone is the invariant sphere of our assumptions!

As a matter of fact the direction-cosines are not invariants, but writing l' , m' , and n' for l , m , and n in equation (2) makes no difference. However, we must state that we are only quoting and trying to explain Einstein.

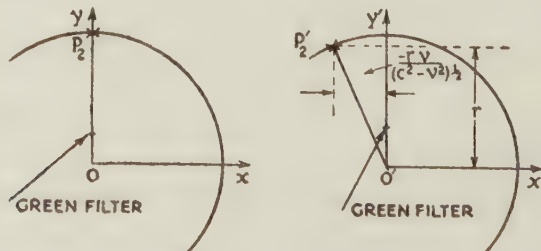
Born transforms the 4-dimensional points $P_1 : [r, 0, 0; r/c]$, $P_2 : [0, r, 0; r/c]$ and $P_3 : [0, 0, r; r/c]$ into $P'_1 : [r\sqrt{(c-v)/(c+v)}, 0, 0; r/c\sqrt{(c-v)/(c+v)}]$,

$P'_2 : [-rv/\sqrt{(c^2 - v^2)}, r, 0; r/c \cdot c/\sqrt{(c^2 - v^2)}]$

and $P'_3 : [-rv/\sqrt{(c^2 - v^2)}, 0, r; r/c \cdot c/\sqrt{(c^2 - v^2)}]$. Events P'_1 and P'_2 (or P'_3) are not simultaneous. And here Born comes to a stop, having proved nothing. He was to have shown that the wave-front is spherical in S' . In fact we can proceed further in the same vein. At time $t' = r/c \cdot c/\sqrt{(c^2 - v^2)}$ the wave-front on the x' axis is at $rc/\sqrt{c^2 - v^2}$. The events $[rc/\sqrt{(c^2 - v^2)}, 0, 0; r/c \cdot c/\sqrt{(c^2 - v^2)}]$,

$$[-rv/\sqrt{(c^2 - v^2)}, r, 0; r/c \cdot c/\sqrt{(c^2 - v^2)}]$$

and $[-rv/\sqrt{(c^2 - v^2)}, 0, r; r/c \cdot c/\sqrt{(c^2 - v^2)}]$ are then simultaneous in S' and the wave-front thus discovered clearly lies on a sphere of radius $\sqrt{[r^2 + r^2 v^2/c^2 - v^2]} = rc/\sqrt{c^2 - v^2}$ at time $r/c \cdot c/\sqrt{(c^2 - v^2)}$. What more could be desired? But let us pause and reflect. Have we transformed points lying on a sphere of S ? Not at all. The point P_1 of S is *not* transformed but another point $P_{11} : [rc/(c-v), 0, 0; r/(c-v)]$ of S . We have obviously transformed three points non-simultaneous in S and lying on an ellipsoid, now in S , into three points simultaneous in S' and lying, as a poor consolation, on a sphere in S' . The situation is somewhat like that of a clown in a circus, who drops as many packets as he picks up and when picking up again, drops the ones he picked up earlier.



There are contradictions galore. Let us look at P_2 and P'_2 . Imagine a tiny, say, green filter placed across the y and y' axes at a distance from the origins $\ll ct$. The space-point $(0, r, 0)$ on the y axis in S receives green light (emitted white) but the

transformed space-point in S' , $(-rv/\sqrt{(c^2 - v^2)}, r, o)$, which does *not* lie on they' axis, is green by transformation but white according to the principle of relativity. The two systems are simply not equivalent. Even the abracadabra of non-simultaneity cannot manage to dispel this contradiction. The 4-dimensional points on the surface $x^2 + y^2 + z^2 - c^2 t^2 = 0$ are undoubtedly transformed by Lorentz equations into *some* points on the surface $x'^2 + y'^2 + z'^2 - c^2 t'^2 = 0$ by the simple device that $x'^2 = x^2 - \phi$ and $c^2 t'^2 = c^2 t^2 - \phi$ where $\phi = (x^2 v^2 / c^2 - 2xvt + v^2 t^2) / (1 - v^2 / c^2)$, but when we depart from these algebraic considerations and go into the physical consequences of these transformations, antinomies appear.

Following the principle of relativity, we took the equations of the expanding wave as $x^2 + y^2 + z^2 = c^2 t^2$ (sphere with origin of light-pulse at O) and $x'^2 + y'^2 + z'^2 = c^2 t'^2$ (sphere with origin of light-pulse at O'). From these equations we derived the Lorentz equations as usual. We then returned to apply these equations to the very phenomenon on which they were based. The spatial co-ordinates of the points of intersection of the wave-front with the three axes of the two systems are:

In $S(x, o, o)$, (o, y, o) and (o, o, z) at time t .

In $S'(x', o, o)$, (o, y', o) and (o, o, z') at time t' .

Mark the second equation. Faithful still (until a contradiction appears) to the principle of relativity, we take the centre of the wave-front in S' at O' and the equation of the wave-front in S' , $x'^2 + y'^2 + z'^2 = c^2 t'^2$, as we first assumed it. For the instant t we put $x = y = z = r = ct$ in S and the distances x' , y' and z' in S' are then found by transformation to be $r\sqrt{(c - v)/(c + v)}$, r and r at time $t' = r/c\sqrt{(c - v)/(c + v)}$. There is no cube to be seen here as Born avers.

We submit that, when there is something wrong, it is usually with our 'primitive' notions to which we may profitably return again and again. We are thankful to Professor Max Born for the stimulus he has provided with his discussion.

Bungalow No. 11

Chopan: Dis. Mirzapur (U.P.)

India

G. H. KESWANI

PROFESSOR DINGLE ON FALSIFIABILITY: A SECOND REJOINDER

IN Professor Dingle's Reply¹ to my earlier Rejoinder to him,² he writes: 'Professor Grünbaum's statement that the Kennedy-Thorndike experiment could falsify the contraction hypothesis is simply incorrect. The experiment gave a null result, which was consistent with the contraction and the "time dilatation" operating together. . . . No conceivable result of the experiment could have falsified the contraction hypothesis, or even increased the probability of its falsity.' And concerning my claim that there is experimental evidence supporting Einstein's postulate that the velocity of

¹ H. Dingle, this *Journal*, 1960, II, 145

² A. Grünbaum, this *Journal*, 1960, II, 143-145

light is independent of the velocity of its source against Ritz's rival hypothesis, Professor Dingle says: 'Professor Grünbaum has apparently not seen a paper in *Mon. Not. R.A.S.*, 1959, **119**, 67. There is at present no experimental evidence at all either for or against this postulate.'

My reasons for now offering a concluding rejoinder to Professor Dingle are the following: (1) I deem it essential to call attention to the error vitiating the conception of falsifiability and confirmability which underlies his rejection of my interpretation of the Kennedy-Thorndike experiment; (2) unlike Professor Dingle, I do attach importance to a correct assessment of the *ad hoc* charge against the Lorentz-Fitzgerald contraction hypothesis. For such logical clarity not only serves to articulate the general logical features of an *ad hoc* hypothesis but also throws light on a significant episode in the pre-history of the Special Theory of Relativity; and (3) Professor Dingle neglected to mention that he himself is the author of the paper cited by him as having invalidated the generally held view, which I had endorsed, that there exists observational evidence cogently supporting Einstein's postulate against that of Ritz.

In order to make quite clear what is at issue between Professor Dingle and myself, I shall show first that attention to an important conditional clause in my original Note could have obviated his erroneous supposition that my interpretation of the Kennedy-Thorndike experiment rests on the illicit neglect of an implicit auxiliary hypothesis concerning the rates of moving clocks. And it will then become apparent that Professor Dingle's criticism turns entirely on dubious principles of inductive reasoning, which he invokes with equanimity as if they were integral to our funded knowledge of scientific methodology.

Professor Dingle overlooked unfortunately that I had stated explicitly that within the framework of the aether theory as modified by the Lorentz-Fitzgerald contraction hypothesis, a *positive* result in the Kennedy-Thorndike experiment is *predicated* on the assumption that 'the period of the light source [as measured by the clocks in the *aether system*] does *not* itself depend upon this velocity [relatively to the aether system]'.¹ But this proviso is tantamount to the recognition that the *positive* outcome of the Kennedy-Thorndike experiment is *predicated* on the original aether-theoretic assumption that the rates of clocks in moving inertial systems are *the same* as those of the clocks in the privileged aether system. To see this, we note first that the *proper* frequency of a moving light source (i.e. the frequency as measured by the clocks of the *moving system*) is *the same* as the frequency which that light source would exhibit when at rest in the aether system. And it is then patent what would be the consequence of assuming a *dependence* of the period of the light source, as measured by the clocks of the *aether system*, on the velocity of that source relatively to the aether by postulating a reduction in the *frequency* of the light in the ratio of $\sqrt{1 - \beta^2}$: 1, as measured by the clocks of the *aether system*. Clearly what follows is the occurrence of the 'time-dilatation' in the sense of the reduction in the rates of moving clocks by a factor of $\sqrt{1 - \beta^2}$ as compared to the clocks of the aether system. As Professor Dingle rightly remarks, the coupling of the assumption of this time dilatation with the Lorentz-Fitzgerald contraction hypothesis does indeed yield a null outcome for the Kennedy-Thorndike experiment: since, as measured by the clocks of the *aether system*, the round-trip times for the vertical and for the *contracted* horizontal arms of that experiment are respectively given by²

¹ A. Grünbaum, this *Journal*, 1959, **10**, 49

² Ibid.

PROFESSOR DINGLE ON FALSIFIABILITY

$$T_v = \frac{2L}{c\sqrt{1-\beta^2}} \quad \text{and} \quad T_h = \frac{2l}{c\sqrt{1-\beta^2}},$$

the corresponding round-trip times registered by the relatively *retarded* clocks of the *moving* system would be *independent* of the velocity of the apparatus with respect to the aether, thereby precluding a shift in the interference fringes due to the changes in that velocity.

The issue between Professor Dingle and me arises therefore *not* from my having been unaware of the fact that the null outcome of the Kennedy-Thorndike experiment was 'consistent with the contraction and the "time dilatation" operating together'. Instead, our disagreement centres on whether this logical fact can be used to show that the null outcome of the Kennedy-Thorndike experiment may *not* be construed as having falsified the Lorentz-Fitzgerald contraction hypothesis, as he claims.

In order to demonstrate the untenability of this thesis of his, we must ask whether the hypothesis that the contraction and the time dilatation operate together is a *methodologically* acceptable interpretation and inductively a legitimate alternative to my construal of the null outcome as having falsified the contraction hypothesis. More specifically, the question before us is which of the following two interpretations of the negative result of the experiment is *methodologically* the more warranted one:

- (i) The Lorentz-Fitzgerald contraction hypothesis is *falsified* in the sense of being (highly) disconfirmed, though not *conclusively* falsified, and this *independent* test of it shows that its espousal in response to the Michelson-Morley experiment was *not ad hoc*, and
- (ii) The compound auxiliary hypothesis comprising *both* the Lorentz-Fitzgerald contraction *and* the time dilatation is *confirmed*.

It is my contention that interpretation (i) is warranted methodologically while (ii) definitely is *not*. Professor Dingle maintains without supporting argument that interpretation (i) is illicit merely because there exists in the purely logical sense an alternative interpretation (ii) *affirming* the Lorentz-Fitzgerald contraction. But this reasoning would be valid, only if he could show—which he cannot—that the proffered rival (ii) will pass muster *methodologically* by *not* being demonstrably *ad hoc*. As I emphasised, however, in my original Note, *unlike* the version of the aether theory containing *only* the modification that there is a Lorentz-Fitzgerald contraction, 'a version of the aether theory incorporating *both* the Lorentz-Fitzgerald contraction *and* the Lorentz-Larmor-Poincaré time dilatation is vulnerable to the *ad hoc* charge.¹ For the assumption that the Lorentz-Fitzgerald contraction *and* the time dilatation operate together does *not in principle* lend itself to any independent test whatever. And it is, of course, precisely this latter fact which provides the justification for *rejecting* the *doubly-amended* variant of the aether theory despite its *consistency* with observational findings and for *accepting* Einstein's theory instead.

It is clear now that Professor Dingle's denial of the falsifiability (disconfirmability) of the Lorentz-Fitzgerald contraction hypothesis by the null outcome of the Kennedy-Thorndike experiment founders on his attempt to preserve the hypothesis from refutation by dint of incorporating it in an augmented auxiliary hypothesis which is indeed *ad hoc*. Accordingly, his criticism has failed to refute my original assessment of the

¹ A. Grünbaum, this *Journal*, 1959, 10, 50

Kennedy-Thorndike experiment as being capable of furnishing an independent test of the contraction hypothesis.

What can be said of the arguments put forward by Professor Dingle in the paper of 1959 cited by him (and in his earlier paper in *Bull. Inst. Phys.*, 1958, 9, 314) in his endeavour to show that extant experimental evidence does *not* lend support to Einstein as against Ritz? I shall refrain from the lengthy rebuttal which his altogether unconvincing theses would require, since I am content to rest my case on the following two refutations of the Ritz hypothesis: (i) the critique given in W. Pauli's careful treatment,¹ and (ii) a very recent report of an experiment by A. M. Bonch-Bruevich published in English translation in *Physics Express*, November 1960, pp. 11-12, which explicitly repudiates Dingle's contention.² Bonch-Bruevich writes: 'a direct experiment showing that the velocity of light is independent of the velocity of motion of the source of emission relative to the observer is meaningful in view of the fundamental significance of the [Einstein] postulate. Dingle discusses this (*Nature*, 1959, 183, 1761) in a recently published comment. From his comment it follows that the author is not aware of the results obtained from an experiment on the direct verification of the second [Einstein] postulate which was performed in 1955.³

'The experiment consisted of comparing the times t_1 and t_2 required to traverse the path $L = 2000$ m by light emitted from two moving sources which were the equatorial limbs of the sun. . . . The results of the experiment seem to us convincing, and their repetition in another form (for example, when using the radiation from excited atoms or ions as the moving source) will probably not be of great interest at the present time.'

ADOLF GRÜNBAUM

University of Pittsburgh
Pittsburgh 13, Pennsylvania

A REPLY TO PROFESSOR GRÜNBAUM'S REJOINDER

THE point at issue is very simple, namely: could the Kennedy-Thorndike experiment have falsified the contraction hypothesis? Two results were possible: (i) a fringe shift; (ii) no fringe shift. In case (i) the most natural explanation would have been contraction, alone or supplemented by some modification, depending on the amount of shift, of atomic frequency. In case (ii) the most natural explanation would have been contraction, supplemented by the particular 'time dilatation' which Lorentz had suggested in 1904.⁴ In neither case, therefore, could the result conceivably have

¹ W. Pauli, *Theory of Relativity*, London, 1958, pp. 5-9

² The Russian original of this paper 'On the Direct Experimental Verification of the Second Postulate of the Special Theory of Relativity' appeared in *Optika i Spektroskopia*, 1960, 9, 134-135. An English translation *Optics and Spectroscopy* is published by the Optical Society of America. I am indebted to Professor A. Janis for this reference to the work of A. M. Bonch-Bruevich.

³ At this point, Bonch-Bruevich cites two earlier papers of his in *Optika i Spektroskopia* for 1956 and 1957 respectively, and another paper of his in *Doklady Akad. Nauk*, 1956, 109, 481.

⁴ *Proc. Amst. Acad.*, 1904, 6, 809

PROFESSOR GRÜNBAUM'S REJOINDER

falsified the contraction hypothesis. It does not matter in the least whether any alternative explanations were possible. In the absence of Einstein's theory (which did not exist when the contraction hypothesis was mooted), these would have been not only possible, but by far the most likely, explanations. Hence the Kennedy-Thorndike experiment could not have falsified the contraction hypothesis.

That is really all there is to say about the matter. The rest of Professor Grünbaum's discussion on this point is mere verbiage.

It might be added that the Lorentz transformation (based on FitzGerald contraction and time dilatation together, as distinct from contraction alone) was an *ad hoc* attempt to save the Maxwell-Lorentz electromagnetic theory, but not an *ad hoc* attempt to explain any particular experiment—certainly not one whose result was published in 1932. It could be falsified by a falsification of the electromagnetic theory, such as would be provided, for example, by an experimental demonstration that the velocity of light is c with respect to its source alone. This was a part of Ritz's hypothesis.

With regard to my paper on that hypothesis, its purpose was to show that 'the critique given in W. Pauli's careful treatment' was inconclusive. I think this is obvious when once it is pointed out, and no-one has attempted to question it. In so far as Professor Grünbaum rests his case on Pauli's treatment, therefore, its situation is precarious. I am indebted to him, however, for calling my attention to Bonch-Bruevich's experiment, of which I had not heard. The use of opposite limbs of the Sun instead of a double star overcomes the defect in de Sitter's test, since the velocity of the sources does not change appreciably during the passage of the light, but from the account in *Physics Express* it is clear that it has no bearing on Ritz's hypothesis because it presupposes what it is held to prove. The times t_1 and t_2 are not measured kinematically, but calculated from the interpretation of the formula $c = n\lambda$ given by the Maxwell-Lorentz theory. But that theory implies that the velocity of light is independent of the motion of the source, so it is illegitimate to use it in this connection. Only a purely kinematical measurement of the times will do, such as the process I suggested in the *Nature* letter which Bonch-Bruevich mentions. I have gone into the whole question of the relations between electromagnetism and kinematics in *Philosophy of Science*, 1960, 27, 233; there I wrote, for example, 'there can scarcely be a doubt that, so long as we remain within the framework of present electromagnetic theory, we shall find complete consistency with the Lorentz transformation, for the electromagnetic equations themselves are invariant to it'. Bonch-Bruevich's experiment merely exemplifies this statement.

HERBERT DINGLE

REVIEWS

CHALLENGE TO DUALISM

MANY a theoretical physicist has toyed with the idea of a unified, comprehensive theory combining cosmology, quantum mechanics, and relativity. However, cosmology as well as quantum physics abound with an increasing amount of intrinsic difficulties, incompatible traits, "messy" facets, conflicting hypotheses and gigantic extrapolations. To cite a few of our present worries:

(a) In cosmology we may choose between relativistic world models (static or expanding, expanding or contracting, or both); the steady-state; the electric universe (positive charge-excess).

(b) In current physical theory we seem to have the choice of macro-causality, microcausality, or acausality. The interaction of elementary particles requires a tentative causality postulate (elimination of infinite constants). Does causality mean that every effect has a cause? That there will be no absorption before emission? Is causal interdependence only an ordering in time? Or only local interaction (field theory)?

(c) Negative probability may not be a nightmare for a pure mathematician, but for the physicist it presents a fountain of unintelligibility.

(d) Bohm's conjecture of a sub-quantum level where $\delta x \delta p < h$ (in analogy to $\delta p \delta x < ma$).

(e) The obscurities concerning renormalization in quantum electrodynamics.

(f) The property of the δ -function that it is infinite at one point (where it is not zero), yet approaches infinity in such a way that its integral is unity.

(g) What is a quantum-mechanical formalism? Is it not a fact that quantum mathematics is already an *interpreted* formalism?

(h) The aphysical so-called functional representation of corpuscles and its relation to the "onde moyenne" (Destouches and Aeschlimann).

This list would be incomplete without reference to the most powerful tenet in quantum theory: the wave-particle duality. It is the merit of one of our foremost quantum theorists, Alfred Landé, to have challenged the validity, the very plausibility of that duality being a fundamental, genuine characteristic of nature.

Landé's *From Dualism to Unity in Quantum Physics*¹ is an exciting, crucial, though highly controversial, piece of incisive reasoning, in part truly original—a rare instance of independent yet responsible probing into some of the "mysteries" in quantum theory.

¹ Cambridge University Press, Cambridge, 1960. Pp. xvi + 114. 18s. 6d.

Landé proposes to show that wave-particle duality is not an immanent property of the micro-realm of matter. He does not deny that complementarity procedure interrelating corpuscular and undulatory phenomena is legitimate and even useful. But he refuses to accept the prevalent viewpoint that wave-model and particle-model can be treated on a par, as if we encountered here two equivalent aspects of certain physical entities. In accord with Born's statistical interpretation of the ψ -function he pleads for a unitary quantum mechanics of particles and he contends that such an exegesis applies equally well to discrete corpuscles as to field oscillators, to the neutron no less than to the vibrating field. "Actually, an electron always behaves exactly as a particle ought to behave. It is a myth that it sometimes misbehaves." Hence it is sheer solecism to conceive of the "real" particle being also a "real" wave: parity between the concrete and an abstraction is muddled reasoning, at least in an exact science.

No doubt, Landé—who presents his case with the acridity of a "young man in anger", with logical rigour and the inexorable spirit of a crusader—anticipated many of the counter-arguments of the adherents to the customary duality standpoint. Thus, he does not question Born's qp -commutation rule, qp -periodicity being wave-like, Schrödinger's p -operator rule, and, most important of all, the amazing wave-like relation between co-ordinates and momenta, de Broglie's quantization $E = h\nu$, $p = h/\lambda$. Neither does he deny the veracity of experimental results. Yet he does not regard quantum rules as mathematical expressions of duality, uncertainty, and complementarity; according to him these rules can be explained in a simple manner by recourse to postulates of invariance, symmetry, the law of unitary transformation, and other similar exact principles; in brief: by invoking non-quantal "axioms".

Could one marshal authoritative support for Landé's claim that the "wavicle" dualism is a recondite concept, alien to common sense, a specious philosophic profundity? Well, de Broglie himself preferred the expression of a wave being *associated* with a corpuscle. And Schrödinger stressed again and again that the wave-picture is neither the observed facts nor nature; indeed, wave-picture and observable facts are not even in one-one correspondence; we must *think* in terms of waves, but the observed pattern manifests itself in the form of single particles. But he also suggested that particles are illusions, generated by high crests of waves in a continuous medium, "explosion-like events within the wave-front": the unitary wave interpretation.

Although I do not wish to indulge in analytic word-spinning, I feel that Landé's insistence upon appraising physical issues exclusively in the light of objective theory cannot conceal the fact that he is forced to argue ontological issues. Electrons and electric fields are real things no less than trees, whereas ψ -waves are mental abstractions? All right, this is a reasonable

viewpoint. But it represents a philosophic position; for instance, Popper with whom Landé is in strong agreement on vital questions concerning philosophy of science regards fields of forces as "metaphysical" and he has given good reasons for his assertion. Landé too is philosophically committed; his approach to realism in physics definitely makes sense. Somehow it appears as if he is wary of any explicit "confession of faith" in order not to violate the simple rules of "orderly thinking" in regard to quantum theory.

It is almost a painful task to have to concur with Landé's conclusions that dualism is based upon a grave epistemological blunder, namely upon equating a thing, for example an electron, with a property (wave-like statistical distribution of many electrons). If I am a man, I need not fear a black cat; if I am a mouse, then an encounter with this cat might be most unpleasant. Likewise, if I accept a unitary particle interpretation, then the wave-attribute will not do me any harm; if I am a dualist, then I must become prey of the 'real' matter-wave! However, when one studies Landé's method of demonstration, one has to admit that he does not just offer fantastic conclusions, but a concatenation of neat grounds why his main thesis can demand approval. Quantum theorists used to approach the particle-wave issue in the form of a *modus ponens*: from $p \supset q$ and p to infer q , but they could not reconcile this inference with the equally valid canon: from $p \supset q$ and not- q to infer not- p (*modus tollens*). This scheme is grossly oversimplified and therefore vulnerable, but it might illustrate an attitude. The gist is that Landé is right in emphasising that the working physicist regards the electron as a discrete particle, that is, as p in the above scheme.

A considerable section of the book is devoted to the *irreducible* statistical character of quantum events. Landé does not treat von Neumann's proof as valid; moreover, he urges that there is no need for such a proof at all, since Born's statistical particle interpretation is self-explanatory and self-sufficient. Von Neumann first evolved a mathematical formalism (?) and later endowed it with a physical meaning. But, Landé avers, any given formalism permits a variety of interpretations. Of course. But von Neumann only proceeded in a manner which is frequently practised; to wit, Green's theorems and functions (e.g. Dirac), Kronecker's δ_{ij} (a purely formal concept, for no such function actually exists), abstract algebra etc., while von Neumann faced a concrete situation and applied a "formal" method to eliminate the *ad hoc* nature of certain traits in quantum theory. I do not see any difference between his approach and that of Carathéodory. In fact, Landé himself, as will be presently shown, could not do without such aids in order to drive his anti-dualistic message home. In parenthesis: I do not concur with the view that von Neumann's proof is circular and thus invalid (or useless); on the contrary, I regard it as a paradigm of that inferential rigour which one can muster in this type of enterprise.

REVIEWS

Landé's armoury for combating duality consists of the postulates of symmetry, invariance, reproducibility, the law of unitary transformation, of cause-effect continuity, and the necessary conformity between statistical fact and *a priori* theory. Let us arbitrarily commence with the law of cause-effect continuity.

Leibniz enunciated three kinds of continuity. It was an ingenious move on Landé's part to invoke that variety which postulates: when cases (antecedents) continually approach and are finally merged in each other, the events (consequents) must do so too. Now, this is a mathematical principle and Landé renders it plausible that it also works in physics (as Leibniz anticipated). This postulate is mobilised to justify the occurrence of cause-effect continuity together with discontinuous acausal events. In other words, between two states M and N are intermediate states "sometimes M" and "sometimes N". In short: there are indeterminate cases governed by statistical averages, where the continuity of cause-effect is *a priori* incompatible with determinism. Landé introduces the notion of "fractional equality" which, owing to Leibniz's principle, is supposed to bridge the dichotomy between equality and inequality, so that $M \sim N$. On this showing, utilizing the Stern-Gerlach "splitting-effect", combining it with the postulates of symmetry and reproducibility regarding a test result, he explains the strange quantum phenomenon that a separator or filter, constructed to pass M-state particles, does in fact pass N-state particles: the N-state particles jump, because they are in contact with the separator, to the new state M and can therefore pass. In analogy, the rejected N-state particles jump to a non-M state so that they surrender their passage tickets. Without adducing Planck's h or Reichenbach's three-valued logic, the forementioned postulates and the splitting effect suffice to show why there are states intermediate between "passed" and "blocked"—states which are sometimes passed, sometimes rejected by the M-instrument (or N-filter, as the case may be). Heisenberg's qp -uncertainty becomes thereby a special case and need not be regarded as a *specific* quantum effect! The symmetry requirement leads to $P(M \rightarrow N) = P(N \rightarrow M)$ so that one arrives at the probability of transition from state M to state N in an N-test and from N to M in an M-test. One has here a statistical counterpart of the two-way symmetry relation, viz. reversibility in classical mechanics (deterministic). The physics involved in Landé's reasoning seems to me sound, but I wish he had chosen a more appropriate term for "fractional equality". The logic of this expression is incondite and vulnerable.

He provides some convincing additional arguments in support of intransigent indeterminism and acausality; Planck's insistence upon statistical regularities being reducible to dynamic, accurate and absolute, that is deterministic, laws holding for the individual event, is cogently repudiated. Whence the conformity between formal inference and individual facts in

statistical ensembles? Must we not admit a "miracle" of co-operation? Landé maintains that this state-of-affairs is irreducible—it is a primitive property of physical nature. That is, if a miracle is an ever recurring experience, then one might as well accept it without bewilderment and christen it, say, Gauss error law. And yet, that there is no causal explanation for the omnipresent correlation between statistical dispersion and random theory, is a corroding issue, if only for deeply entrenched psychological habits—an observation readily granted by Landé. I think that the search for sufficient causes for every event is indeed foredoomed, *provided* one is resigned to the idea that some essential features of quantum mechanics are definitely ultimate.

When dealing with the metric of probabilities, he makes use of a so-called law of unitary transformation which he equates with the law of interference of probabilities in quantum theory. He proposes to furnish rigorous proof combined with sound physical postulates

why the statistical manifestations of particles should obey a law of "interference of probabilities" *via* a complex-imaginary amplitude called Ψ .

And also, why co-ordinates and momenta of particles can be depicted in a periodic wave-like relation. This topic suffers from a too condensed treatment. Further, there is no such "law", but a beautiful theorem or rule in linear algebra applying to hermitian and quadratic forms. Linear transformations A which possess the property $G(Au, v) = G(u, Av)$ (let $G(u, u)$ be the primitive form), lead, under strictly defined conditions, to linear transformations U which leave invariant the fundamental hermitian form $G(u, u)$; those U transformations are designated as *unitary*, or in the real case, as *orthogonal*. Again, under special conditions we can obtain a complete orthogonal system of vectors, which can be normalized, rotated, etc. We are interested in the fact that the matrix is unitary and that for linear mappings any algebraic relation between matrices is invariant under certain transformations.

There is an old story about a Russian peasant who visited Moscow for the first time and landed at the Zoo where he noticed the giraffes. "Look, Mother," he whispered to his wife, "look what the Bolsheviks have done to our horses". Well, I felt like this peasant when I read Landé's interpretation of magic squares, unitary transformation, and metric of geometrical structures. However, on very patient scrutiny it must be conceded that his derivation of P-correlation is a beautiful and flawlessly reasoned piece of work, in its elegant deductive procedure a true model for analytic thinking in theoretical physics. I only wish he had not compressed his arguments in such a tight fashion: the "unpacking" of the diverse claims is at times true labour. If P is the probability of transition and P -tables are appropriately conceived of as unit magic squares, then by correct application of unitary transformation and the assumption that ψ is a matrix ($P = |\psi|^2$), one must

arrive at an expression demonstrating with mathematical necessity why we have interference of probability amplitudes ψ . Landé is unduly worried about unitary transformation being the only possible correlation law linking P-matrices; a uniqueness proof for the monopoly of unitary transformation in regard to P-correlation law might probably be attainable, on sheer symmetry (or group-theoretic) grounds. Still, as long as the validity of approach proposed by Landé is not disproved, his arguments hold on mathematical and physical evidence. I suggest that equation (2b) on page 45 be altered so that it contains the explicit mention of *unit matrix* to avoid possible confusion.

Perhaps the most valuable and consequential insight gained by Landé concerns a feasible, unambiguous interpretation of the Schrödinger wave equation. One sometimes wonders whether the ψ -function will ever exhaust the creative (often truly poetical) minds of imaginative physicists and philosophers. That poor function or wave is supposed to occupy a definite region in space; lacks location even in fictional space; is a pilot wave; a quantum potential; guides events; gradually expands and occasionally shrinks suddenly; contains no physics at all, etc., etc. Do you remember what that suave gentleman replied when reprimanded for having made some derogatory remarks about a lady-friend? "Granted, there is a lot to say in her favour, but the 'other' is more interesting." Indeed, some flagrant absurdities appealed to physicists for various reasons and Landé fulfilled a necessary "therapeutic" task by eschewing the mere *interesting* readings into the ψ -function.

The wave equation is not a deterministic equation of motion; $\psi(r, t)$ does *not* represent a physical state in space varying with time; Bohm is right when he stresses that the Schrödinger equation depends upon the representation one uses; however, it is misleading to depict the ψ -function in three-dimensional space; hence this kind of "model" ought to be excised from our picture-store in quantum mechanics. It is the mathematical formalism of quantum mechanics that determines, through the square of the modulus of the complex function $\psi(q)$, the probability that a measurement on the system (particle) will produce the values of the co-ordinates to be in the element dq of *configuration* space. The normalization condition for ψ -functions, $\int |\psi|^2 dq = 1$, follows when we integrate $|\psi|^2$ over all configuration space, for we lay down, by definition so to speak, that the sum of all the possible probability values of the system's co-ordinates should be equal to 1. Landé proves, adopting a different approach, that a complex imaginary function $\psi(q, p)$ is the periodic probability amplitude function $\psi(q, p) = \text{const} \exp(2i\pi qp/h)$ and is equivalent to the Schrödinger rule and the Born commutation rule; Heisenberg's uncertainty relations (holding for every pair of canonical conjugate variables, not only for q and p , but also for E and t etc.,) can easily be obtained from the exponential form $\exp(2i\pi qp/h)$

and $\exp(2i\pi Et/h)$ for $\psi(q, p)$ and $\psi(E, t)$ respectively. Landé's list of six current views as to the "correct" interpretation of the wave-function could become a collector's item displaying inconsistency in reasoning among men of science.

Schrödinger had, at some stage, uttered his doubts as to his unitary wave interpretation and also whether ψ -waves have physical existence of their own. But he died irreconciled to Born's statistical interpretation and thus to the notion of transition probability. He believed that the latest findings conform neither to the corpuscular nor to the wave theory—an observation which is reflecting our present ignorance regarding interaction of elementary particles; but he applied his scepticism to customary dualism itself, not only to quantum field theory. That neither the particle nor the wave "picture" can claim truth was his final agnosticism. Landé is right: it can be demonstrated that the Schrödinger equation is a probability metric; after all, a curve is not a "real" thing. Jeans' "waves of knowledge", the recourse to the observer's "state", the cryptic allusions to "material waves"—all these startling expressions are examined by Landé and found wanting. The wave is the statistical "disposition" of the particle, and wave-like phenomena are the direct consequences of non-quantal postulates.

I fully concur with Landé when he pinpoints the defects of a unitary wave conception; it could never account for corpuscular phenomena. On the other hand, Born's statistical particle theory permits us to explain the periodic wave-like relation between co-ordinates and momenta. I remember how Schrödinger enchanted us with the simile of the violin and its sound to illustrate the particle-wave distinction. It made sense, at that time, but Weizsaecker's dualism of two "pictures" which end in an abstract formula can be shown to be an "emergency exit"—the product of a facile positivistic defeatism. How come, one asks, that Born himself did not pronounce his original unitary theory with the same zest and power of conviction of Landé, all wave and duality talk notwithstanding? That "real" things are no more than *Gestalten* or invariant constants must lead to mathematical idealism, to a flight into a-physical realms. Is it possible that Born has always been right in his attempt at deriving quantum mechanics from classical probability considerations? In other words, has Landé proved that the Born of yesterday has to be protected against the Born of today? Before we can answer this question, let us glance at the obvious reason for Born's reorientation towards Bohr, Heisenberg, etc. No doubt, it is quantum field theory (relativistic) and the intractability of bosons (e.g. photons, mass ~ 0 , spin 1). Photons are not localized and should not be treated as particles at all. Now, bosons obey the Einstein-Bose statistics with symmetric state-functions and cannot be subjected to Boltzmann's classical statistics; in contrast the Fermi-Dirac statistics (i.e. Pauli principle) with anti-symmetric state-functions applies to fermions, e.g. electrons, with spin $\frac{1}{2}$. In the

latter case too is Boltzmann statistics without value. Schwinger stipulated that fields with integer spin *must* obey commutation relations and fields with half-integer spin anti-commutation relations. (See also Lüders' theorem). Since the quantum theory of fields seems to permit the possibility of different types of fields, the symmetry principle cannot be used as a master key any longer. Can we reasonably assume that a field may possess different ψ -functions from that of the form $\exp(2i\pi qp/h)$ which Landé selected? I think that his general formulation of the wave-function cannot be invalidated, *provided* one employs a notation such as $\psi(p, q)$, $\psi(K, q)$, $\psi(E, q)$ in a consistent manner; we need the general expression for the operator p and all the diverse functions $\psi(G, q)$, for any arbitrary operator $G(p, q)$ can then be computed. Born is concerned about the grave fact that quantum theory does not suffice to explain the field properties of fundamental particles and about the question whether conservation of energy will still obtain in the new theory, which may or may not be quantized. However, Landé emphasises that he does not wish to deal with this big problem of relativistic field theory. His goal is the deduction of quantum mechanics from non-quantal propositions and a sharp definition of the wave-function. One could put it this way: he confines himself to intrapolation within quantum mechanics proper and shuns bold extrapolations. This self-imposed restriction renders Landé's case more convincing even to a carping mind. The issue at stake: can we afford to ignore relativistic quantum field theory when discussing the particle-wave duality? I think his assertion that this is possible has been vindicated by his arguments.

As to be expected, Landé cannot accept the Copenhagen interpretation of quantum theory. Any subjectivism in physics is anathema to him and it is interesting to note that Schrödinger, in one of his last papers, also roundly condemns the language of Copenhagen epistemology. I do not wish to dampen Schrödinger's and Landé's polemic spirit in this matter, but it seems to me as if the opponents of that interpretation have not taken the trouble to study the so-called Copenhagen commitment in detail. N. R. Hanson's attempt to remedy this omission did us a salutary service by showing that only a cautious analysis of that viewpoint can keep the discussion free from misapprehensions and inappropriate claims. Bohr and Heisenberg could persuasively argue that their ideas need not result in a physics of solipsism, if one plays the game according to their set of rules! In brief: Landé is right in questioning the validity of duality being an intrinsic feature of nature, but he can hardly deny the descriptive value of that term within a clearly defined frame.

When did duality as a tantalisingly inadequate expression give its first performance? The common opinion is that de Broglie "injected" wave-particle dualism into theoretical physics. Landé corrects this wrong assumption: Einstein's hypothesis concerning photons and Duane's corpuscular

REVIEWS

theory of X-ray diffraction paved the way for the wavicle hybrid. However, current knowledge will compel us to revise our appraisal of those two classic contributions, so that the spurious foundations for duality will be placed in proper perspective.

Voltaire remarked that the secret of being tiresome is to tell everything. To my mind, Landé is aware of the myriad ramifications of his unitary particle interpretation, but he had to concentrate upon some essential aspects and propound them in such a manner that the core became lucid. Will his radically developed Born theory imprecate imminent doom on "wavists" or "waviclists"? I can only hope so, since I believe in the self-correcting process of scientific method. Landé is no iconoclast—he belongs rather to that rare type of creative scientists who engage in excoriating the arguments of co-workers until their theoretical edifices tumble down. Planck, Einstein, Bohr, de Broglie, Born, Heisenberg, Schrödinger, Jordan, Landé—they all "did" physics, they were creative in their respective domains, and then they pondered over the physical, the epistemological significance of their discoveries and later on physical theory in general. The logic, the method responsible for the formulation of physical propositions is often alien to them. Landé is less guilty of that marked shortcoming.

I found the book under discussion a fascinating piece of powerful argumentation, a true requiem for duality, if not a solemn one, then at least an honest one. There are some typographical errors, but they do not become irritating; still I wish that a new edition would display at some detail intermediary steps in order to spare the reader the bother of searching for necessary links. The discussion on reproducibility does only in part support Landé's main thesis; nor is the thorny problem of the discernible individuality (?) of a particle satisfactorily dealt with. Yet the analysis of the meaning of the term "state" in quantum physics is an example of precise and relevant elucidation. Similarly, Landé succeeds in illuminating the most important implications of the Second Law, consequences often ignored or wrongly presented. Here too one cannot impute to him "wisdom of hindsight", as he had enunciated his views long before they became either fashionable or neglected; and also here is the affinity with Born evident in all vital aspects.

The saga of duality is recounted wherever theoretical or nuclear physicists pursue their studies, their teaching, their research. I hope, I urge, that Landé's book will be read by them, argued *after* his arguments have been listened to. Landé is no Don Quixote . . . and the castles of duality are not even "real" windmills.

WOLFGANG YOURGRAU

Smith College
Northampton, Mass.

The Biological Way of Thought. By Morton Beckner.

Columbia University Press, New York; Oxford University Press, London, 1959. Pp. viii + 200. 48s.

THIS book is a welcome addition to the classics of biological methodology. It is a worthy successor after thirty years to J. H. Woodger's *Biological Principles* to which it owes much as a pattern and starting point. Not least in this connection is its main aim which is to examine the ways in which concept formation and explanation in the biological sciences differs from those in the other natural sciences. The author has drawn widely from other writers in the philosophy of science and the work might be regarded as a restatement of *Biological Principles* with the benefit of the writings of contemporary logicians and methodologists upon which to draw. It is perhaps a pity that one writer which the author has ignored is Popper. However, there is much that is original in the book, and the clarity of writing and the precision of many formulations of ideas and problems makes the book lucid, thought provoking and a distinguished contribution to the subject.

The chapter on the Types of Biological Concept is a key to the subject matter of the whole book and much of what follows later depends on the validity of the ideas in this section. The author distinguishes between concepts which can be well-defined (W-defined) and those which can also be effectively defined (E-defined) and this distinction is skilfully used in subsequent discussion. The conditions imposed on a W-definition are intended to be an application of the verifiability criterion of meaning. They are that

1. A test-procedure is describable for determining whether an entity is a member of the extension of the concept being defined.
2. The test-procedure consists of a finite number of logical and/or physical operations.
3. Each operation of the test-procedure is compatible with the laws of nature and of logic.

A concept is E-defined with respect to a non-empty class K of entities if it is W-defined, if K is equal to or is included in the extension of the concept, and a test-procedure is describable for deciding with a high degree of probability that an entity is a member of K, and if this test-procedure, known to be compatible with the laws of nature can be carried out at present. When the class K and the extension of the concept coincide the concept is E-defined. The concept of 'gene locus' is W-defined, but not E-defined. The author gives the concept of 'prospective potency' in an embryo as being E-defined but not W-defined. It is not easy to see how this can be from the definition of the E-defining relation where a W-definition appears to be an essential requirement of the E-definition.

From these notions we find that there are three types of concept in biology which in one sense are fully unique in biological theory.

1. Polytypic concepts:

This class of concepts is defined with reference to a set of properties $f_1, f_2, f_3 \dots f_n$ so that each individual in the class possesses a large number of the properties, each f is possessed by large numbers of the individuals and no f is possessed by every individual. This is not the same as the Boolean product of all the properties, and the author discussed the significance of the difference. It appears that this vagueness in the polytypic concept is its value. Such concepts have important uses in any discussion involving systematics and in the analysis of other types of explanation which are discussed later in the book.

2. Historical concepts:

Criteria are given for their use which, briefly, indicate that a process is always involved, and that the process is describable as a temporal sequence of stages. The term 'hybrid corn' is an example of an historical concept.

3. Functional concepts:

'A concept C is "functional" if, and only if, C is W -defined (...) such that every application of C to a system s and process F is (logically) false unless there exists a system s' and a process F' (other than s and F) and at least one environment and state of s' in which s' does exhibit F' and in which s' would not exhibit F' if (1) F failed to occur, or (2) s is not a part of s' or that environment.' The names of most organs are functional concepts together with the names of responses and so forth.

Before discussing the rôle that these concepts play in biological theorising, the author discusses different types of model found in biological theory. After distinguishing 'model = interpretation of calculus', the 'mathematical model' and 'replica = model' Professor Beckner draws attention to a fourth type, an example of which he finds in the work of Rashevsky. Other examples are the well-known biophysical models for cell division, the evolutionary models of Fisher, Sewall Wright, etc., the neural models of McCulloch and the feed-back models of the cyberneticists. They are a technique of explaining or quasi-explaining a class of phenomena (called the field) by references to properties of systems (called the subject) which are either parts of the system exhibiting the field phenomena or of its environment.

I do not find it easy to see the point of the introduction of this fourth kind of model or why it should be called a model at all. Its nature appears to be that of an explanation of events at one level in terms of parts of a lower microscopic level. Beckner points out that there are two assumptions characteristic of his models. These assumptions are, firstly, law-like statements about the subject of the model and, secondly, existential statements about the existence of elements of the subject which simplify because

they serve to reduce the structural complexity of the field. This seems to me to be characteristic of any change of explanatory level. The whole tenure of the argument becomes apparent when one realises that Beckner is accepting Hempel and Oppenheim and Braithwaite's requirements that for them an explanans must be true. By adopting this criterion of the satisfactoriness of an explanation, he wishes to use the concept of model in which the explanans may be known to be false or certainly not known to be true. He then talks of his models as being quasi-explanations. Much of what he says about levels of generality and the reducibility of theories has been said before in, I feel, clearer terms. It is a pity in this context that he nowhere refers to Woodger's description of the way in which theories have levels, which is found in the early parts of *Biology and Language*. Woodger's description is, to me, much clearer, and nowhere conceives that the statements of the higher level are necessarily true. Indeed, if this were the case, we should never progress in science by providing alternative hypotheses.

I have described the keystones of Professor Beckner's views in some detail because I hope it can be seen from my account how important these topics are for any analysis of biological problems such as are found in Systematic Theory, Evolutionary Theory, or in Organismic Biology. It is impossible to review in detail the analysis he offers in these various fields, and I shall only comment to give some indication of the approach he adopts. On one trivial point of terminology I think it is a pity he has used the term genetic analysis in the sense of historical analysis. It is used clearly and explicitly with this meaning as opposed to the technical biological sense. Philosophers and psychologists will be quite familiar with this usage; on the other hand, students of the biological sciences will not, and with a well-developed science of the same name at their finger tips, will find the book confusing until they have become accustomed to this other usage.

In both the chapters on Systematics and Genetic Analysis polytypic concepts play an important rôle. One is left wondering whether the apparent success of the concept is as real as it appears. It is the sort of concept for which one feels intuitive sympathy, but the proof of its success is in its use in statements. I would recommend that a detailed instance of its usefulness be worked out. It is easy to use the concept for an intuitive analysis and to feel that 'in principle' it should be satisfactory. However, its logic is only discoverable by taking a piece of evolutionary theory and seeing whether statements involving the concept generate the sort of consequence that is desired. It is a pity that the author did not do this in the book.

The next part of the book is a discussion of the rôle of functional concepts in teleological explanations. The writings of Braithwaite, Sommerhoff, and Nagel are discussed. In particular Beckner points out that Sommerhoff's analysis in terms of directive correlation always involves the success of the

REVIEWS

object in reaching its goal. Any teleological explanation needs to give an account of goal-seeking behaviour in those cases where the system is only partially successful—of how the object gets onto the target even if it does not score a bull.

The later chapter on selection theory and organismic biology involve a synthesis of the use of functional and historical concepts with model-explanations. The methodology involved in part-whole relationships is discussed quite fruitfully. A detailed discussion of the points raised would require another book. I am sure that in departments where biological methodology is taken seriously this book will generate ideas for many years to come.

R. F. J. WITHERS

Isaac Newton's Papers and Letters on Natural Philosophy and Related Documents.

Edited with a General Introduction by I. Bernard Cohen assisted by Robert E. Schofield.

Cambridge University Press. 1958. Pp. 501. 70s.

In this book Professor Cohen has assembled all Newton's published papers and letters relating to optics, chemistry, and other branches of natural philosophy which were issued either during his lifetime or shortly after his death. This volume may therefore take its place on our bookshelves beside the *Principia* and the *Opticks*, for the three together can be regarded as 'the complete corpus of Newton's writings on physical science that actually influenced the scientists and thinking men of the "Age of Newton"'. In addition to works from Newton's own pen, the editor has included Halley's review of the *Principia* and his account of Newton's theory of the tides which he wrote for the benefit of James the Second, Fontenelle's obituary notice (in English) of Newton and the Seventh and Eighth Boyle Lectures delivered by Richard Bentley in 1692, being 'A Confutation of Atheism from the Origin and Frame of the World'. The book ends with a useful Appendix of references to Newton in Birch's *History of the Royal Society*. The different sections are prefaced by contributions from several American scholars: Thomas S. Kuhn on the optical papers, Marie Boas on the chemical papers, Perry Miller on Bentley and Newton, Robert E. Schofield on Halley and the *Principia*, and C. C. Gillispie on Fontenelle and Newton.

Nearly half the book is devoted to optics. Almost all the papers on this subject were originally published in the *Philosophical Transactions*. As Dr Kuhn points out in his introduction to these papers, Newton was the

first great scientist to use the then new medium of the journal article to give an account of a new branch of experimental science. Moreover, it was his work that first stimulated international debate in a scientific journal. These papers and Dr Kuhn's discussion make it clear that the essential difference between Newton and other contemporary students of optics was his great success in interpreting his experiments mathematically. On the other hand, his skill as a practical designer of optical instruments was not particularly remarkable, and indeed he was often wrong in his suggestions. His celebrated reflecting telescope remained essentially a curious toy for half a century, until James Hadley succeeded in grinding a parabolic mirror.

Newton's achievements as a chemist are assessed by Marie Boas who emphasises that a full understanding must still await a careful analysis of the extant chemical notebooks in the University Library, Cambridge. Despite the scores of alchemical works in his library, Newton's approach to chemical problems was close to that of Boyle who had no use for mysticism in science. Underlying Newton's chemistry, no less than his physics, was the corpuscular hypothesis of matter. But, although the 'mechanical philosophy' endeavoured to explain all chemical and physical properties of bodies solely in terms of geometry and the motion of particles, Newton—unlike Descartes—felt obliged to retain one ostensibly occult influence which he at first called 'sociability' and later 'attraction'. Nevertheless, he tried, as far as possible, to eliminate its occult overtones. His notorious disavowal of hypotheses was influenced by his sensitivity on this score.

For the student of the philosophy of science, Newton's correspondence with Bentley and its lucid analysis by Perry Miller will be of particular interest. The contrast between the great literary scholar and the even greater natural philosopher is nicely pointed in the Third Letter where Newton remarks that 'Those things which Men understand by improper and contradictory Phrases, may be sometimes really in Nature without any Contradiction at all.'

A word of praise must be given to Bernard Cohen for his urbane General Introduction in which he argues that the papers and letters assembled in this volume clarify one of the major Newtonian problems: why did Newton maintain that he was formulating no hypothesis concerning the mechanism of gravitation, and yet at the same time refer (in the General Scholium at the end of the Third Book of the *Principia* in the second edition of 1713) to the possible rôle of the elastic aether as the agent of gravitational attraction? In the early optical papers, in the famous letter to Boyle and in a long statement read many years before to the Royal Society, Newton repeatedly discussed this hypothesis. Cohen claims that, although Newton refused to admit an *untestable* hypothesis into the framework of the *Principia*, he nevertheless wished to remind the cognoscenti that he had framed long ago an elegant hypothesis that might account for gravitation and indeed for other forces of nature.

REVIEWS

All those concerned must be warmly commended for producing so useful and timely a volume. Some readers may regret the decision to reproduce each document in facsimile from the original publication, in order to present to the modern scholar the 'very works' studied during Newton's lifetime and in the decades after his death. This inevitably makes the book harder to read and gives it an uneven appearance. But this is a minor objection to so welcome an addition to the literature on Newton.

G. J. WHITROW

The Evolution of Living Things. By H. Graham Cannon, F.R.S.
Manchester University Press, 1958. Pp. x + 180. 12s. 6d.

PROFESSOR Graham Cannon has addressed this book to the layman because he despairs of influencing professional biologists. In it he develops three main themes. The first of these is concerned to show that the views of Lamarck have been so misrepresented by biologists as to mask their explanatory value. In the second, explanations of animal evolution based on natural selection and on genetical theory alone are shown to be unsatisfactory in that they cannot, in principle, provide a complete explanation. The third argues that the influence of the environment on organisms is inherited and that this does adequately explain those aspects of evolution for which the author believes natural selection to be inadequate.

The task is approached through a brief historical sketch covering the six thousand years preceding this century. The brevity of this introduction is achieved by a very considerable simplification of the views and concepts described. In the five ensuing chapters the historical and biological information necessary to the first two themes are described with wholly admirable clarity and economy. The field covered is broad. It includes the genetics of singly and multi-factorial inheritance, cytology, natural selection, the inheritance of environmentally influenced characteristics, experimental embryology and finally comparative anatomy and biochemistry. The seventh, and final, chapter discusses evidence in support of Graham Cannon's thesis that Lamarckian explanations are essential for evolutionary theory.

The criticisms of Neo-Mendelian genetics and Natural Selection advanced are familiar but are re-stated with an impelling and admirable lucidity, although in covering so wide a field for an audience innocent of biological concepts it has been necessary to make considerable omissions. One of these omissions is of crucial importance to the development of the third theme. The argument of this last theme involves a lemma in which an attempt is made to show that evolution must have occurred before living things developed the cellular mechanism on which Neo-Mendelian genetics depends. The reader is led to this conclusion through an introduction to the nucleus, to chromosomes, and to genes and is told that living things of the

present must have possessed ancestors devoid of chromosomes. It is held that in consequence of this Neo-Mendelian genetics is not an explanation for the evolution of such ancestors. It might be the case that the conclusion is accurate but to arrive at it on the information given is an over-simplification. Neo-Mendelian explanations of inheritance have now moved from the area of cytology to that of biochemistry. It is not the case that living things are known to exist now that do not contain nucleic acid and it is believed to be the presence of this substance that is essential for Neo-Mendelian explanations. Had the argument been based on the present universality of nucleic acid then on the argument presented it would have been necessary to infer the presence of this substance in ancestral organisms and hence the possibility of Neo-Mendelian mechanisms for them.

The theoretical framework in which the arguments are developed is less clear than the admirable accounts of biological data. Throughout the thesis there is the implicit assumption that teleological explanations are valid in biology. Unfortunately there is nowhere an adequate refutation of the view clearly expressed (by authors such as Nagel, in Feigl and Brodbeck, *Readings in the Philosophy of Science* 1953) that teleological explanations play no valid part in biological explanation. Lamarck's laws are invoked as being fundamental for the argument without the part they play in the theory being made clear. These laws are discussed as if they were explanatory hypotheses but the manner in which they are stated is of the form of generalisations of observation records. For example, Lamarck's first law of 1809 is stated as 'The development and effectiveness of organs are proportional to the use of those organs' (p. 60). This states a functional relationship similar to Ohm's law and suggests no mechanism by which organs might appear anew that does not involve teleology.

In the argument developed in his final chapter Professor Graham Cannon appears to accept the hypothetico-deductive view of scientific method, at least in part. He tells his reader that the principle he postulates has always been put forward in the past as an idea which cannot be tested experimentally. He then interprets his argument in terms of our current physico-chemical concepts. Unfortunately his interpretation requires us to go back to the origins of life, and this we cannot, in principle or in fact do.

F. T. C. HARRIS

Theories of Scientific Method: The Renaissance through the Nineteenth Century.

By Ralph M. Blake, Curt J. Ducasse, and Edward H. Madden. Edited by Edward H. Madden.

University of Washington Press, Seattle, 1960. Pp. iv + 346. \$6.50.

PHILOSOPHY of science is a comparative newcomer among the branches of philosophy, and some would say that it has not yet, or even cannot ever,

become wholly autonomous. In any case it has not yet its canon of classic texts. One therefore approaches this volume with interest to see how the selection has been made. The Preface announces that 'Our authors are primarily concerned with these persistently recurring philosophical topics and problems: theory construction, hypothesis, causality, lawfulness, probability, the experimental methods, the uniformity of nature and the rule of succession (the problem of induction), the universality of causality, and the problem of discriminating the a priori and the empirical elements of science'. The book consists of thirteen chapters, each devoted to a single author, or a closely related group of authors, and each provides an exposition of the author's views with copious quotations.

The first two chapters deal generally with discussions of scientific method in the Renaissance, including those of Leonardo and the sixteenth-century astronomers. These are followed by four seventeenth-century writers: Bacon, Descartes, Hobbes, Newton; then by a brief account of Hume's view of causation; and finally by six nineteenth-century writers: Herschel, Whewell, Mill, Jevons, Peirce and Chauncey Wright. This selection leaves one with the somewhat depressing reflection that, at least to the end of the nineteenth century, those who are known mainly for their writings on scientific method are not the most exciting of philosophers. In the whole list represented here, apart from the brief acknowledgment of Hume, the only philosopher in the first rank is Descartes. Is scientific method really so unrewarding as not to attract the best philosophical minds? The question has only to be asked to betray an important shortcoming in the selection: where are the other members of the apostolic succession of the history of philosophy—Locke, Leibniz, Berkeley, and Kant? Has the textbook stereotype of their philosophical positions so overlaid their actual work that no one now reads them to discover what they have to say on the topics with which this book deals? And yet it might be argued that they are more worth reading for their views on theory-construction, hypothesis, and the relation between the a priori and empirical, than most of their nineteenth-century successors. Without them this survey cannot claim to be complete, and the omission is more regrettable since, in spite of all that has been written on these philosophers, good commentaries on their philosophy of science in its contemporary scientific setting are not easily available.

This book would not therefore be a complete introduction to the 'classic sources', but for those authors who are included it does a workman-like job. The earlier chapters convey something of the polemics of a period when philosophy of science was the manifesto of the new scientific movement rather than the analysis of established methods. Two problems are introduced: the question of *certainty* and the question of *reality*. We are shown how the Renaissance humanists rejected natural science on the ground that it cannot attain certain truth, and how Descartes and Hobbes, accepting this

as a criterion for the value of science, attempted to show that truth in science is attainable and to derive it from self-evident principles. Huygens and Newton are credited with being the first fully to abandon the view that science can be absolutely certain and final. The problem of the real or fictional nature of hypotheses was confused throughout this period with the problem of certainty, and is not always carefully distinguished in some of the comments of the authors of this book, especially in the chapter on Descartes. But the chapter on the Renaissance Astronomers gives an interesting account of the interactions of truth and falsity, and reality and fiction, in their claims for the planetary hypothesis, and in their generalisations of these views from astronomy to physics. Clavius and Kepler both gave comparatively clear accounts of hypotheses as potentially *real* descriptions of the system of the world, having high probability in relation to the evidence, but which may nevertheless be *false*. The problem of certainty hardly arises again after the end of the seventeenth century (although the contribution of Locke on this point might have been included), but the problem of fictional hypotheses does continue to arise, and it is a pity that it is not pursued into the writings of the nineteenth-century positivists.

The chapters on Bacon and Descartes do overdue justice to both these philosophers in showing how in each case their own scientific investigations forced them to modify their initial philosophical positions, even at the cost of inconsistency. Thus Bacon the empiricist and compiler of facts is shown to have in practice a sound idea of the function of hypotheses, and Descartes the rationalist spends much time analysing the role of experience in science. There is a kind of homeostasis about the new scientific method which tends to stability in spite of the exaggerations on one side or the other of its principal exponents, and, as always in a progressive period of philosophy, inconsistency is less important than creative novelty.

Newton's view of hypothesis is a well-worn subject, but it is treated here with detail, balance, and freshness. This account, and also the chapter on Descartes, suggests that our understanding of seventeenth-century philosophy of science would benefit from a study of the use of the terms 'deduction' and 'induction'. Descartes uses the words almost synonymously in the *Regulae*, and Newton uses such startling locutions as 'deduced from the phenomena'. I have never seen a careful examination of this matter, although with regard to Newton I have a private hunch that he was tempted to generalise the genuine *deduction* of the law of gravitation from Kepler's laws (which he calls *phenomena*), together with his Rules of Reasoning, which are very general assumptions applicable to all phenomena. This reverses the order of deduction of the usual hypothetico-deductive form, and since Newton carries it out in the *Principia* with detailed cross-references to the Rules of Reasoning at each stage, it is plausible to suppose that he thought such a method would be universally applicable to 'higher-level' theories.

REVIEWS

He was not proved wrong until the acceptance of Maxwell's theory, and in the meantime Ampère had made a bold attempt to construct electromagnetic theory on the same pattern.

The interest of the nineteenth-century chapters largely centres in the juxtaposition of Herschel, Whewell, and Mill, and the consequent deflation both of Mill's originality (Bacon's inductive methods were expanded and transmitted to him by Herschel), and of his insight into the nature of theories (much more adequately discussed by Whewell). The chapters on Jevons and Peirce introduce opposing views on the nature of probable inference and its relation to the justification of hypotheses.

In the case of a few of the authors studied here one feels after reading this book that one may well leave it at that—they have historic interest but little to say to twentieth-century philosophy of science. But in other cases one is stimulated to return to the originals, guided by the sign-posts to be found here, and that perhaps is not the least valuable function of the book.

MARY B. HESSE

ANNOUNCEMENT

MEETINGS OF THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

1960

- 3 October: Professor G. L. S. Shackle, 'Time, Nature, and Decision'
- 7 November: Dr S. Vajda, 'The Building of Models and the Theory of Games'
- 5 December: Mr G. Buchdahl, 'Theories and Models'

1961

- 16 January: Presidential Address, Professor K. R. Popper, 'The Growth of Knowledge'
- 13 February: Professor E. Gombrich, 'Theory and Observation in the Work of Leonardo da Vinci'
- 13 March: Professor K. Menger, 'A Counterpart to Occam's Razor in Pure and Applied Mathematics'
- 17 April: Mr P. Alexander, 'Science and Epistemological Atomism'
- 15 May: Dr E. W. Bastin 'Self-Organising Mechanisms as Models for Scientific Theories'
- 12 June: Professor M. Munitz, 'The Logic of Cosmology'